

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 06:52:09 ; Search time 1311 Seconds
(without alignments)
11162.167 Million cell updates/sec

Title: US-10-616-263-29

Perfect score: 2472

Sequence: 1 tatgagccttcggaactgtg.....aaaaaaaaaaaaaaaaaaaaa 2472

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 segs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_16Dec04.*

1: Geneseqn1980s.*

2: Geneseqn1990s.*

3: Geneseqn2000s.*

4: Geneseqn2001as.*

5: Geneseqn2001bs.*

6: Geneseqn2002as.*

7: Geneseqn2002bs.*

8: Geneseqn2003as.*

9: Geneseqn2003bs.*

10: Geneseqn2003cs.*

11: Geneseqn2003ds.*

12: Geneseqn2004as.*

13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2472	100.0	2472	3	AAAL6632 Human sec
2	2472	99.8	2472	6	Abk35878 cDNA sequ
3	2406.6	97.4	2572	6	Abq54852 Human ova
4	2399.2	97.1	2776	3	AAa08567 Human cha
5	1961.6	79.4	2562	5	AAa83823 DNA encod
6	1763.2	71.3	1921	3	AAa99959 Human sec
7	1503.8	60.8	1536	3	AAa77414 Human ORF
8	1314.2	53.2	1546	13	ADP43721 Human PMM
9	1253.8	50.7	1274	3	AAa99977 Human sec
10	593	24.0	694	6	Abk35896 cDNA sequ
11	422.4	17.1	458	2	AAa97580 Extended
12	422.4	17.1	458	12	ADP18847 Human sec
13	413.4	16.7	421	3	AAa77867 cDNA enco
14	413.4	16.7	421	4	AAI28605 Colon tum
15	413.4	16.7	421	8	ABZ32791 Human col
16	352	14.2	619	10	ADD71001 Human Mck
17	344.6	13.9	348	3	AAa02899 Human sec
18	327.8	13.3	371	2	AAa86645 EST clone
19	269	10.9	332	4	AAa58781 cDNA #145
20	254.8	10.3	350	2	AAa26699 Human gen

21	244.2	9.9	268	3	AAA41503	Aaa41503 Human sec
22	186.4	7.5	313	2	AAa41298	Aax41298 Human sec
23	186	7.5	233	3	AAA45883	Aaa45883 Human sec
24	130.4	5.3	248	2	AAa41081	Aax41081 Human sec
25	123	5.0	134	2	AAH85281	Aah85281 Human sin
26	123	5.0	134	2	AAH85282	Aah85282 Human sin
27	113.4	4.6	407	13	ADQ57549	Adq57549 Novel can
28	95.6	3.9	385	5	ADL37765	Adl37765 Human ova
29	95.6	3.9	385	5	ADL37765	Adl37765 Human ova
30	95.4	3.9	2136	11	ACN92757	Acn92757 Breast ca
31	94.6	3.8	534	13	ACN58719	Acn58719 Cotton gy
32	93.6	3.8	546	13	ADR62275	Adr62275 Cotton cd
33	93	3.8	349	5	ADL44149	Adl44149 Human ova
34	93	3.8	483	5	ABV59155	Abv59155 Human pro
35	93	3.8	1767	12	ADO62272	Ado62272 Transcrip
36	92.4	3.7	469	13	ACN62049	Acn62049 Cotton gy
37	91.8	3.7	381	13	ACN55708	Acn55708 Cotton an
38	91.8	3.7	487	13	ACN48060	Acn48060 Cotton pr
39	91.6	3.7	3420	12	ADH61306	Adh61306 INTSIG en
40	91.4	3.7	390	5	ADL43918	Adl43918 Human ova
41	91.4	3.7	1762	6	AAL45656	Aal45656 Human can
42	91.4	3.7	1762	6	AAL45657	Aal45657 Human can
43	91.2	3.7	318	13	ACN51559	Acn51559 Cotton an
44	91.2	3.7	549	13	ACN52320	Acn52320 Cotton an
45	91	3.7	510	13	ADR64316	Adr64316 Cotton cd

ALIGNMENTS

RESULT 1

AAAL6632

ID AAAL6632 standard; cDNA; 2472 BP.

XX

AC AAAL6632;

XX

DT 16-JUN-2000 (first entry)

XX

DE Human secreted protein clone pp314_19 nucleotide sequence SEQ ID NO:29.

XX

KW Human; secreted protein; immunostimulant; immunosuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.

OS Homo sapiens.

XX

PN WO200009552-A1.

XX

PD 24-FEB-2000.

XX

PF 13-AUG-1999; 99WO-US018298.

XX

PR 14-AUG-1998; 98US-0096622P.

PR 17-AUG-1998; 98US-0096815P.

PR 04-SEP-1998; 98US-0099229P.

PR 23-OCT-1998; 98US-0105168P.

PR 08-JAN-1999; 99US-0115234P.

PR 12-FEB-1999; 99US-0119931P.

PR 18-FEB-1999; 99US-0120575P.

PR 30-APR-1999; 99US-0132020P.

PR 11-AUG-1999; 99US-0148424P.

XX (GENY) GENETICS INST INC.

PA

XX Jacobs K, McCooy JM, Lavallie ER, Collins-Racie LA, Evans C;

PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;

PI Wong GG, Clark HF, Fechtel K;
XX WPI; 2000-205979/18.
DR P-PSDB; AAY94912.
XX
PT New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing, hematopoiesis
PT regulating, tissue growth, activin/inhibin antiinflammatory or tumor
PT inhibition activity.
XX
PS Claim 38; Page 498-499; 641bp; English.
XX
XX AAA16618 to AAA16697 encode the human secreted proteins given in AAY94898
CC to AAY94980, isolated from human adult brain, adult thyroid, adult
CC retina, foetal carcinoma, adult blood, adult neural, foetal kidney, adult
CC placenta, adult testis, whole embryo, adult cartilage, kidney, foetal
CC brain, adult thymus, foetal placenta, adult uterus, adult tumour, and
CC adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans and
CC animals. The polynucleotides can be used as markers for tissues in which
CC the protein is preferentially expressed, as molecular weight markers on
CC Southern gels, and as chromosome markers or tags to identify chromosomes
CC or to map gene positions. The proteins can be used in the treatment of
CC immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention
XX
SQ Sequence 2472 BP; 796 A; 475 C; 502 G; 699 T; 0 U; 0 Other;

Query Match 100.0%; Score 2472; DB 3; Length 2472;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGAGCCTTCGAACTTGGGAGAGACTACAAAGTTTGGTTGTTATGGTCCCTTAGT 60
DB 1 TATGAGCCTTCGAACTTGGGAGAGACTACAAAGTTTGGTTGTTATGGTCCCTTAGT 60
QY 61 TGGGCTCATACATTTGGGGTGGTACAGAAATCAAAGCAGCCCTGTGTTCCAAATACCTAA 120
DB 61 TGGGCTCATACATTTGGGGTGGTACAGAAATCAAAGCAGCCCTGTGTTCCAAATACCTAA 120
QY 121 AAACGACGATTCCTGAGCAGATAGTCTGGGACTTTCAAATCTTTCAGAGAGCCAAAT 180
DB 121 AAACGACGATTCCTGAGCAGATAGTCTGGGACTTTCAAATCTTTCAGAGAGCCAAAT 180
QY 181 CAGGGGGAAGTAGCAGGCTTGCATCTTCAGGTAAAGAGCAGCTTTGAACTCGAGCTTC 240
DB 181 CAGGGGGAAGTAGCAGGCTTGCATCTTCAGGTAAAGAGCAGCTTTGAACTCGAGCTTC 240
QY 241 ATATCGAAGAGAGATGAAATACACAGTTGGATTAGAAAGAACTGGCTTCTTGTAGCT 300
DB 241 ATATCGAAGAGAGATGAAATACACAGTTGGATTAGAAAGAACTGGCTTCTTGTAGCT 300
QY 301 GGGATATCTTTTCATAGGTGTCATCTTGGAAACATACCTTTTTCAGAGGTCGCAAGCAG 360
DB 301 GGGATATCTTTTCATAGGTGTCATCTTGGAAACATACCTTTTTCAGAGGTCGCAAGCAG 360
QY 361 TCTGTAATATTTTCAGTCTCAAACCAACAAAGAGTATTGAAGAGTGAAGTAAATAAT 420
DB 361 TCTGTAATATTTTCAGTCTCAAACCAACAAAGAGTATTGAAGAGTGAAGTAAATAAT 420
QY 421 ATTTGGAATTTACATTTTGTCAATTAATCATTTCTATGCTGATTAGCTTCATAAACATTGA 480
DB 421 ATTTGGAATTTACATTTTGTCAATTAATCATTTCTATGCTGATTAGCTTCATAAACATTGA 480

DB 421 ATTTGGAATTTACATTTTGTCAATTAATCATTTCTATGCTGATTAGCTTCATAAACATTGA 480
QY 481 ACTTTTGTGATTTTATAGCCACAATGCTGCAATATTCTTAACTTAACTTCTTAAAGATAATT 540
DB 481 ACTTTTGTGATTTTATAGCCACAATGCTGCAATATTCTTAACTTAACTTCTTAAAGATAATT 540
QY 541 TTTAATGTTTAAACCGTGATAATGCAATAATAAGAAAATGCTGGTTTACAAAATAAAACG 600
DB 541 TTTAATGTTTAAACCGTGATAATGCAATAATAAGAAAATGCTGGTTTACAAAATAAAACG 600
QY 601 GTCTTCACCTAGTTACACCTGAAGTAAGATGCTCGTTTGGGAAGCTGAAGGCCATCAT 660
DB 601 GTCTTCACCTAGTTACACCTGAAGTAAGATGCTCGTTTGGGAAGCTGAAGGCCATCAT 660
QY 661 GTGTGAAGAGTGAACCACTGACAACTGAGAGAGTCAGGACCAACATTTTGTCTTGAAG 720
DB 661 GTGTGAAGAGTGAACCACTGACAACTGAGAGAGTCAGGACCAACATTTTGTCTTGAAG 720
QY 721 AATTGTACATCATGCTATGCGCCCTCAGGTAGGCTGAGCAGCTGCACAATGGCTTGG 780
DB 721 AATTGTACATCATGCTATGCGCCCTCAGGTAGGCTGAGCAGCTGCACAATGGCTTGG 780
QY 781 AGGTTACGCTGTGACAACTCAGTCCCTCAGCTCTGCTCAGTCACTCTTTTGGTCAACA 840
DB 781 AGGTTACGCTGTGACAACTCAGTCCCTCAGCTCTGCTCAGTCACTCTTTTGGTCAACA 840
QY 841 TCCCAATTTTAAAGATCCTGACAGCCTCCATACAGAAATCATGTGTCAAAGCTTCAGT 900
DB 841 TCCCAATTTTAAAGATCCTGACAGCCTCCATACAGAAATCATGTGTCAAAGCTTCAGT 900
QY 901 TGGCTTATTTACAGCTGTTCTTTGCTGCAACCTGATGGAATGTTTTCAGAGATTAGGCT 960
DB 901 TGGCTTATTTACAGCTGTTCTTTTGTGCAACCTGATGGAATGTTTTCAGAGATTAGGCT 960
QY 961 GACACCCACCACTGCTATTAGATTAATAAATAAATCTTTTGTAGCTCTTTTGCATCAGT 1020
DB 961 GACACCCACCACTGCTATTAGATTAATAAATAAATCTTTTGTAGCTCTTTTGCATCAGT 1020
QY 1021 CAACTGCTGATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
DB 1021 CAACTGCTGATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
QY 1081 TTTGTTTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
DB 1081 TTTGTTTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
QY 1141 AGAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
DB 1141 AGAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
QY 1201 AGGCAACATCATTTTGAAGAGATTTAATTGCTGCTTAAAGGTCGCTGCTGCTGCTGCTGCT 1260
DB 1201 AGGCAACATCATTTTGAAGAGATTTAATTGCTGCTTAAAGGTCGCTGCTGCTGCTGCTGCT 1260
QY 1261 TTTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
DB 1261 TTTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
QY 1321 ACCTATCAAAAATCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
DB 1321 ACCTATCAAAAATCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
QY 1381 CACTTCTGACACTGGAGAGGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
DB 1381 CACTTCTGACACTGGAGAGGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
QY 1441 AGCTTGGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
DB 1441 AGCTTGGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
QY 1501 CCGTGGCCAAAAGTTATACATCCATCTTTTGAAGCAGTTCTCAATGATGCTGCTGCTGCTGCT 1560
DB 1501 CCGTGGCCAAAAGTTATACATCCATCTTTTGAAGCAGTTCTCAATGATGCTGCTGCTGCTGCT 1560

QY 1561 TGCATAGACAGAAATGGAGTGAATCTGTGATGGAAACCCCTGACTAAATGACGGAACACA 1620
DB |||||
QY 1561 TGCATAGACAGAAATGGAGTGAATCTGTGATGGAAACCCCTGACTAAATGACGGAACACA 1620
DB |||||
QY 1621 GCCTATTGGATCCCTAGGCTCAATATGCTCTAATAGTATGGAAGTGTGAAGATGTGTG 1680
DB |||||
QY 1621 GCCTATTGGATCCCTAGGCTCAATATGCTCTAATAGTATGGAAGTGTGAAGATGTGTG 1680
DB |||||
QY 1681 CACTGCAAAATTTGGCTCCCAACAAATTTTTCATCTTATTCCTTAATGAAGCAACAATCTG 1740
DB |||||
QY 1681 CACTGCAAAATTTGGCTCCCAACAAATTTTTCATCTTATTCCTTAATGAAGCAACAATCTG 1740
DB |||||
QY 1741 CAGCTTGCTTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCAAGTGTCA 1800
DB |||||
QY 1741 CAGCTTGCTTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCAAGTGTCA 1800
DB |||||
QY 1801 GAGGGCACTGCATGCTCTGAGTTTACACTCAAGGAACCATGGGCTTTGTTGGAGGTGG 1860
DB |||||
QY 1801 GAGGGCACTGCATGCTCTGAGTTTACACTCAAGGAACCATGGGCTTTGTTGGAGGTGG 1860
DB |||||
QY 1861 CTGTACTGAAACTCATTTTGGCTGCATATATCAGACACAGACTCACAAGCCCAAG 1920
DB |||||
QY 1861 CTGTACTGAAACTCATTTTGGCTGCATATATCAGACACAGACTCACAAGCCCAAG 1920
DB |||||
QY 1921 CATTTCTCAAGAGATGAAATGTACTCAAAACAGAACTTCAATTAATTGCTGAACATTTTG 1980
DB |||||
QY 1921 CATTTCTCAAGAGATGAAATGTACTCAAAACAGAACTTCAATTAATTGCTGAACATTTTG 1980
DB |||||
QY 1981 CAGTGGCTTAGAATCTGTTTGGCTCTTTAGAACATGATGAGAGTGAATTTCTCACTGA 2040
DB |||||
QY 1981 CAGTGGCTTAGAATCTGTTTGGCTCTTTAGAACATGATGAGAGTGAATTTCTCACTGA 2040
DB |||||
QY 2041 CATGAAGTATGGACACTTTTGGTTCAGTTTCAAGGAGATTTCTCCCTGTGTTGCTAACTGGCC 2100
DB |||||
QY 2041 CATGAAGTATGGACACTTTTGGTTCAGTTTCAAGGAGATTTCTCCCTGTGTTGCTAACTGGCC 2100
DB |||||
QY 2101 AGATTTGCTTTCACAGTGTGGCTGTGATTTATCAATAGCCAGGAAGAACTCAACTGGTC 2160
DB |||||
QY 2101 AGATTTGCTTTCACAGTGTGGCTGTGATTTATCAATAGCCAGGAAGAACTCAACTGGTC 2160
DB |||||
QY 2161 TTTCTTAAGAGACACAGCTGCTCAATTTGTGACAAAGCTGCTTCCATGAACTGT 2220
DB |||||
QY 2161 TTTCTTAAGAGACACAGCTGCTCAATTTGTGACAAAGCTGCTTCCATGAACTGT 2220
DB |||||
QY 2221 GGGCTCAGCCAGCAACTGACCTTGGACTGTTTGTGACAAAGCTTGTGAGCTCAGCT 2280
DB |||||
QY 2221 GGGCTCAGCCAGCAACTGACCTTGGACTGTTTGTGACAAAGCTTGTGAGCTCAGCT 2280
DB |||||
QY 2281 GGGCTTAGAGACAGCCAAATTTGATTTGGGATCTTTTCATATGTTTGAAGATAAAACTA 2340
DB |||||
QY 2281 GGGCTTAGAGACAGCCAAATTTGATTTGGGATCTTTTCATATGTTTGAAGATAAAACTA 2340
DB |||||
QY 2341 AGAAGATAGCATGTTTGTATTAACAGAGAAACAAATTAATAGTCTGTGTTGGCAATTTGAGA 2400
DB |||||
QY 2341 AGAAGATAGCATGTTTGTATTAACAGAGAAACAAATTAATAGTCTGTGTTGGCAATTTGAGA 2400
DB |||||
QY 2401 AA 2460
DB |||||
QY 2401 AA 2460
DB |||||
QY 2461 AAAAAAAAAAAAAA 2472
DB |||||
QY 2461 AAAAAAAAAAAAAA 2472
DB |||||

RESULT 2
ABK35878/c
ID ABK35878 standard; cDNA; 2472 BP.
XX
AC ABK35878;
XX
DT 08-MAY-2002 (first entry)

XX
DE cDNA sequence #269 encoding novel human secreted protein.
XX
KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;
KW immune deficiency disorder; blood disorder; inflammatory disorder;
KW infectious disorder; allergic condition; neurodegenerative disorder;
KW liver fibrosis; coagulation disorder; gene therapy; antimicrobial;
KW tumour; cancer; hepatotropic; immunosuppressive; antirheumatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200177289-A2.
XX
PD 18-OCT-2001.
XX
PF 29-MAR-2001; 2001WO-US010232.
XX
PR 06-APR-2000; 2000US-0195605P.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCooy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Bowman MR, Spaulding V, Wong GG;
PI Clark HF, Fechtel K, Howes SH, Resnick RJ, Gulukota K, Graham JR;
XX
WPI; 2002-179322/23.
XX
XX Six hundred and twenty three polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for treating
PT immune deficiencies and disorders such as autoimmune disorders.
XX
PS Claim 1; Page 236-237; 393pp; English.
XX
CC The present invention relates to the isolation of novel cDNA sequences
CC which encode human secreted proteins. The cDNA sequences have been
CC derived from a variety of human tissues. The invention also provides a
CC method for producing proteins from these polynucleotide sequences. The
CC proteins are useful for identifying compounds that modulate their
CC activity and production. The sequences of the invention are useful for
CC treating diseases such as hyperproliferative disorders (e.g. cancer),
CC immune deficiency disorders (e.g. severe combined immunodeficiency
CC (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders
CC (e.g. thrombocytopaenia), inflammatory disorders (e.g. arthritis),
CC infectious disorders (e.g. hepatitis), allergic conditions (e.g. asthma),
CC neurodegenerative disorders (e.g. Alzheimer's disease), liver fibrosis,
CC coagulation disorders (e.g. haemophilia), and tumours. The polynucleotide
CC sequences of the invention are also useful in gene therapy. ABK35610-
CC ABK36232 represent the cDNA sequences of the invention that encode for
CC novel human secreted proteins
XX
SQ Sequence 2472 BP; 699 A; 502 C; 474 G; 797 T; 0 U; 0 Other;

Query Match 99.8%; Score 2467.2; DB 6; Length 2472;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2469; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TATGAGCCTTCGGAACCTTGTGGAGAGACTACAAAGTTTTTGGTTGTTATGTCCTTCTTACT 60
DB 2472 TATGAGCCTTCGGAACCTTGTGGAGAGACTACAAAGTTTTTGGTTGTTATGTCCTTCTTACT 2413

QY 61 TGGGCTCATACATTTTGGGTTGGTACAGAAATCAAAAGCAGCCCTGTTTTTCCAAATACCTAA 120
DB 2412 TGGGCTCATACATTTTGGGTTGGTACAGAAATCAAAAGCAGCCCTGTTTTTCCAAATACCTAA 2353

QY 121 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTTCAAAATCTTCAGAAAGGCAAAAT 180
DB 2352 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTTCAAAATCTTCAGAAAGGCAAAAT 2293

QY 181 CCAGGGGAGTAGCAGGCTTGCATCTTTCAGGTAAAGAGCAGCTTTGATCTGAGCTTC 240
DB 2292 CCAGGGGAGTAGCAGGCTTGCATCTTTCAGGTAAAGAGCAGCTTTGATCTGAGCTTC 2233

QY 241 ATATCGAAAGAGAGATGAAAAATACCAGTTGGATTAGAAAGAACTGGCTTCTTGTAGCT 300

|||||
2232 ATATCGAAGAGAGATGAAAATAACAGTTGGATTAGAAAGAACTGGCTTCTTTAGCT 2173
QY
301 GGGATATCTTTTCATAGGTGTCATCTTTGGAAACATACCTTTTTCAGAGGTCTGCAAAAGCAG 360
Db
2172 GGGATATCTTTTCATAGGTGTCATCTTTGGAAACATACCTTTTTCAGAGGTCTGCAAAAGCAG 2113
QY
361 TCTGTAAATTTTCAGTCTCAAGCAAAACAAAGAGATTTGAAGAGTGAAGTAAATTAAT 420
Db
2112 TCTGTAAATTTTCAGTCTCAAGCAAAACAAAGAGATTTGAAGAGTGAAGTAAATTAAT 2053
QY
421 ATTTGGAAATTTTGTGATTTAAATCATTTCTATGCTGATTAGCTTTCATAAACATTGA 480
Db
2052 ATTTGGAAATTTTGTGATTTAAATCATTTCTATGCTGATTAGCTTTCATAAACATTGA 1993
QY
481 ACTTTTGTATTTTATAGCCCAATGCTGCATATTTTCATACCTTTAAATTTCTTAAAGAAATTA 540
Db
1992 ACTTTTGTATTTTATAGCCCAATGCTGCATATTTTCATACCTTTAAATTTCTTAAAGAAATTA 1933
QY
541 TTTTAATGTTTAAACGTTGATTAATGCAATAATAGAAAAATGTGGTTTACAAAATAAAAAAG 600
Db
1932 TTTTAATGTTTAAACGTTGATTAATGCAATAATAGAAAAATGTGGTTTACAAAATAAAAAAG 1873
QY
601 GTCTTCACTAGTTTACCACTGAAAGTAAGATGTCTCGTTTGGAGCTAAGAGCCATCAT 660
Db
1872 GTCTTCACTAGTTTACCACTGAAAGTAAGATGTCTCGTTTGGAGCTAAGAGCCATCAT 1813
QY
661 GTGTAAAGAGTGAACCACTGACAACTGAGAGAGTCAGGACCAACATTTTCTGTGTAAGAG 720
Db
1812 GTGTAAAGAGTGAACCACTGACAACTGAGAGAGTCAGGACCAACATTTTCTGTGTAAGAG 1753
QY
721 AATTGTAACATCATGCTATGGCCCCCTCAGTAGGCTGAAGCAGCTGCACAATGGCTTTGG 780
Db
1752 AATTGTAACATCATGCTATGGCCCCCTCAGTAGGCTGAAGCAGCTGCACAATGGCTTTGG 1693
QY
781 AGGTTACGTGTGTACAACTTCACAGTCTCAGTCTCTGCTCAGTCACCTTTTGGTCACACA 840
Db
1692 AGGTTACGTGTGTACAACTTCACAGTCTCAGTCTCTGCTCAGTCACCTTTTGGTCACACA 1633
QY
841 TCCCATTTTAAAGATCCTGACAGCCCTCCATPACAGAAATCATGTGTCAAGCTTCAGTGATTG 900
Db
1632 TCCCATTTTAAAGATCCTGACAGCCCTCCATPACAGAAATCATGTGTCAAGCTTCAGTGATTG 1573
QY
901 TGGCTTATTCACAGCTATTTCTTGTGCAACCTGATTGAAATGTTTCAAGATTAGGCTT 960
Db
1572 TGGCTTATTCACAGCTATTTCTTGTGCAACCTGATTGAAATGTTTCAAGATTAGGCTT 1513
QY
961 GACACCCACCATGTCTATTAGATTAAATAAACATCTTTTGGATCTTTTGCATCAGTTATCT 1020
Db
1512 GACACCCACCATGTCTATTAGATTAAATAAACATCTTTTGGATCTTTTGCATCAGTTATCT 1453
QY
1021 CAAGTCTGATACCTGTGGTTGTGCAATCCAGTGGACTTTTAGTGTGATCTCAGATCCCTCT 1080
Db
1452 CAAGTCTGAGACCTGTGGTTGTGCAATCCAGTGGACTTTTAGTGTGATCTCAGATCCCTCT 1393
QY
1081 TTGCTTGTGGTGTAGTATATTAACAAGTAAACCTGCTGTATGCTCACCAGAAAGAAAC 1140
Db
1392 TTGCTTGTGGTGTAGTATATTAACAAGTAAACCTGCTGTATGCTCACCAGAAAGAAAC 1333
QY
1141 AGAGCATGTCAAGTGTCTGATCTCTGAGAGCTTTTGTGCTTACAAATTTCCAGAAAATGCTGA 1200
Db
1332 AGAGCATGTCAAGTGTCTGATCTCTGAGAGCTTTTGTGCTTACAAATTTCCAGAAAATGCTGA 1273
QY
1201 AGGCCATCATTTTATAGGAAAGATTAAATGTGACCTTTTAAAGGTCAAAGATTATAGA 1260
Db
1272 AGGCCATCATTTTATAGGAAAGATTAAATGTGACCTTTTAAAGGTCAAAGATTATAGA 1213
QY
1261 TTCCATGTATTACCTGGGATCTCATTTGAAATGTGAGATTCAATTAATGAGGCTATT 1320
Db
1212 TTCCATGTATTACCTGGGATCTCATTTGAAATGTGAGATTCAATTAATGAGGCTATT 1153
QY
1321 ACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTTATCCCGAGA 1380
|||||

Db
1152 ACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGA 1093
QY
1381 CACTTCTGACACTGGAGAAAGAACTGTGTGTGTCTAGTTATGGGGTTTCTCTTTGAAAAATGC 1440
Db
1092 CACTTCTGACACTGGAGAAAGAACTGTGTGTGTCTAGTTATGGGGTTTCTCTTTGAAAAATGC 1033
QY
1441 AGCTTTGGACCAAGCTGCTTAACTTAGGAGGAGCAGCTTAATCAGTGCACCAAGTATGTTCTGT 1500
Db
1032 AGCTTTGGACCAAGCTGCTTAACTTAGGAGGAGCAGCTTAATCAGTGCACCAAGTATGTTCTGT 973
QY
1501 CCGTGTGCCAAAAAGTTATACATCCCATCTTTTGAAGCAGATTTCTCAATATGCAATCGTATTAT 1560
Db
972 CCGTGTGCCAAAAAGTTATACATCCCATCTTTTGAAGCAGATTTCTCAATATGCAATCGTATTAT 913
QY
1561 TGCCTATGACAGAAATTTGGAGTGACTCTGTGATGAAACCCCTGACTAAAAATGACAGGAACA 1620
Db
912 TGCCTATGACAGAAATTTGGAGTGACTCTGTGATGAAACCCCTGACTAAAAATGACAGGAACA 853
QY
1621 GCCTATTTGGATCCCTAGGCTCAATATGCTCAATATGCTCAATATGTTATGGAAGTGTGAAAGATGTGTG 1680
Db
852 GCCTATTTGGATCCCTAGGCTCAATATGCTCAATATGCTCAATATGTTATGGAAGTGTGAAAGATGTGTG 793
QY
1681 CACTGCAAAATTTGGCTCCCAACATTTTCTTCTTATTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1740
Db
792 CACTGCAAAATTTGGCTCCCAACATTTTCTTCTTATTTCTTCTTCTTCTTCTTCTTCTTCTTCT 733
QY
1741 CAGCTTGTCTTCTGTGCAACAGAAATGACACTGCTGCTGGGATGAGCTGAAGCTCACGTGTCA 1800
Db
732 CAGCTTGTCTTCTGTGCAACAGAAATGACACTGCTGCTGGGATGAGCTGAAGCTCACGTGTCA 673
QY
1801 GACGCACTGCATGTCTGCACTTAACTCAAGGAAACCATGGCTTTTGTGGAGGTGG 1860
Db
672 GACGCACTGCATGTCTGCACTTAACTCAAGGAAACCATGGCTTTTGTGGAGGTGG 613
QY
1861 CTGTACTGAAATCTCATTTTGGCTGCATATATACAGACACAAGACTCACAGCCAGCAAGAAAG 1920
Db
612 CTGTACTGAAATCTCATTTTGGCTGCATATATACAGACACAAGACTCACAGCCAGCAAGAAAG 553
QY
1921 CATTTCTCAAAGATGATGATGATCTCAAAACAGAACTTCAATTAATTTGCTGAAGCAATTTTG 1980
Db
552 CATTTCTCAAAGATGATGATGATCTCAAAACAGAACTTCAATTAATTTGCTGAAGCAATTTTG 493
QY
1981 CAGTGCCTTAGAATCTGTGTGTGGCTCTTTAGAAACATGATGAGGTGAATTTCTCACTCA 2040
Db
492 CAGTGCCTTAGAATCTGTGTGTGGCTCTTTAGAAACATGATGAGGTGAATTTCTCACTCA 433
QY
2041 CATGAAGTATGACACCTTTTGTGCTCAGTTCAGGCAGATTTCTCCTGTGTGCTTAACTGGCC 2100
Db
432 CATGAAGTATGACACCTTTTGTGCTCAGTTCAGGCAGATTTCTCCTGTGTGCTTAACTGGCC 373
QY
2101 AGATTGTCTTTCACAGTGTGGCTGTGGATTATACAAATGACAGGAAGAACTCAACTGTGTC 2160
Db
372 AGATTGTCTTTCACAGTGTGGCTGTGGATTATACAAATGACAGGAAGAACTCAACTGTGTC 313
QY
2161 TTTTCTTAAAGACACAGCTGCTCCATTGTGTGCAACAAAGCTGCCCTTCCACATGAAGCTGT 2220
Db
312 TTTTCTTAAAGACACAGCTGCTCCATTGTGTGCAACAAAGCTGCCCTTCCACATGAAGCTGT 253
QY
2221 GGGCTCAGCAGCAACCTTGACCTTTGGACTGTTTGTGCTGCAAAAGCTTTAGTGGCTACAGGT 2280
Db
252 GGGCTCAGCAGCAACCTTGACCTTTGGACTGTTTGTGCTGCAAAAGCTTTAGTGGCTACAGGT 193
QY
2281 GGCTGTAGAGACAGCAATTTGATTTGGATCTTTTCATATGTTATTTGAAGATAAAAACTA 2340
Db
192 GGCTGTAGAGACAGCAATTTGATTTGGATCTTTTCATATGTTATTTGAAGATAAAAACTA 133
QY
2341 AGAAGATGATGTTTGTGTTTAAAGAGAAACAAATAAACTAGTCTGTGTGGCAATTTGAGA 2400
Db
132 AGAAGATGATGTTTGTGTTTAAAGAGAAACAAATAAACTAGTCTGTGTGGCAATTTGAGA 73
QY
2401 AAAAAAAG 2460
Db
72 AAAAAAAG 13

Qy 2461 AAAAAAAAAAAAA 2472:
 Db 12 AAAAAAAAAAAAA 1

RESULT 3
 ABQ54852
 ID ABQ54852 standard; cDNA; 2572 BP.
 XX
 AC ABQ54852;
 XX
 DT 22-AUG-2002 (first entry)
 XX
 DE Human ovarian antigen HMAEL73 cDNA, SEQ ID NO:732.
 XX
 KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive; gene; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200200677-A1.
 PN
 XX
 XX 03-JAN-2002.
 XX
 XX 07-JUN-2001; 2001WO-US018569.
 XX
 XX 07-JUN-2000; 2000US-0209467P.
 PR
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 XX Birse CE, Rosen CA;
 PI
 XX
 XX WPI; 2002-147878/19.
 DR
 XX P-PSDB; ABP41775.
 XX
 XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
 PT cancer), immune disorders, cardiovascular disorders and neurological
 PT diseases.
 XX
 XX Claim 1; SEQ ID NO 732; 2922pp; English.
 PS
 XX
 XX The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56303), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovary and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the

CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents cDNA encoding a human ovarian antigen of the
 CC invention. Note: the sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIFO at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 2572 BP; 762 A; 512 C; 566 G; 729 T; 0 U; 3 Other;

Query Match 97.4%; Score 2406.6; DB 6; Length 2572;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2409; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TATGAGCCTTCGGAACTTGTGAGAGACTACAAAGTTTGGTGTGTTATGTCCTCTTAGT 60
 Db |||||
 Qy 126 TATGAGCCTTCGGAACTTGTGAGAGACTACAAAGTTTGGTGTGTTATGTCCTCTTAGT 185
 Db |||||
 Qy 61 TGGGCTCATACATTTTGGGTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 120
 Db |||||
 Qy 186 TGGGCTCATACATTTTGGGTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 245
 Db |||||
 Qy 121 AAACGACGACATTCCTGACCAAGATAGTCTGGACCTTTCAAATCTTCAGAAGAGCCAAAT 180
 Db |||||
 Qy 246 AAACGACGACATTCCTGACCAAGATAGTCTGGACCTTTCAAATCTTCAGAAGAGCCAAAT 305
 Db |||||
 Qy 181 CCAGGGGAAGTAGCAGGCTTGCAATCTTCAGGTAAAGAGCAGCTTTGTAATCTGAGCTTC 240
 Db |||||
 Qy 306 CCAGGGGAAGTAGCAGGCTTGCAATCTTCAGGTAAAGAGCAGCTTTGTAATCTGAGCTTC 365
 Db |||||
 Qy 241 ATATCGAAAGAGAGATGAAATATACAGTTTGGATTTAGAAAGAACTGGCTTCTTGTAGCT 300
 Db |||||
 Qy 366 ATATCGAAAGAGAGATGAAATATACAGTTTGGATTTAGAAAGAACTGGCTTCTTGTAGCT 425
 Db |||||
 Qy 301 GGGATATCTTTTATAGGTGTCATCTTGGAAACATATCTTTTGCAGAGGTCTGCAAGCAG 360
 Db |||||
 Qy 426 GGGATATCTTTTATAGGTGTCATCTTGGAAACATATCTTTTGCAGAGGTCTGCAAGCAG 485
 Db |||||
 Qy 361 TCTGTAAATTTTTCAGTCTCAAGCAAGCAAAAGAGTATTGAGAGTGAAGTAAATTAAT 420
 Db |||||
 Qy 486 TCTGTAAATTTTTCAGTCTCAAGCAAGCAAAAGAGTATTGAGAGTGAAGTAAATTAAT 545
 Db |||||
 Qy 421 ATTTGGAATTTTACTAATTTTGTCAATTAATCATTTCTATGCTGATTAGCTTCATAAATTCGA 480
 Db |||||
 Qy 546 ATTTGGAATTTTACTAATTTTGTCAATTAATCATTTCTATGCTGATTAGCTTCATAAATTCGA 605
 Db |||||
 Qy 481 ACTTTTGAATTTTATAGCCACAATGCTGCATATTCATATCTTTAAATCTCTTAAAGATTAAT 540
 Db |||||
 Qy 606 ACTTTTGAATTTTATAGCCACAATGCTGCATATTCATATCTTTAAATCTCTTAAAGATTAAT 665
 Db |||||
 Qy 541 TTTTAATGTTAAACGCTGATTAATGCAATTAATAGAAATATGCTTTTACAAATTAACACG 600
 Db |||||
 Qy 666 TTTTAATGTTAAACGCTGATTAATGCAATTAATAGAAATATGCTTTTACAAATTAACACG 725
 Db |||||
 Qy 601 GTCTTCACCTAGTTTACCACTGAAGTATGCTCGTTTGGAGAGCTTGAAGAGCCATCAT 660
 Db |||||
 Qy 726 GTCTTCACCTAGTTTACCACTGAAGTATGCTCGTTTGGAGAGCTTGAAGAGCCATCAT 785
 Db |||||
 Qy 661 GTGTAGAGTGAACCACTGACAACTGAGAGAGTCAAGGACCACACATTTTGTCTTGAAGAG 720
 Db |||||
 Qy 786 GTGTAGAGTGAACCACTGACAACTGAGAGAGTCAAGGACCACACATTTTGTCTTGAAGAG 845
 Db |||||
 Qy 721 AATTGTACATCATCTATATGCCCCCTCAGGTAGGCTGAGAGAGCTGCACAATGGCTTTGG 780
 Db |||||
 Qy 846 AATTGTACATCATCTATATGCCCCCTCAGGTAGGCTGAGAGAGCTGCACAATGGCTTTGG 905
 Db |||||
 Qy 781 AGGTTACGTGTGTACAACTCAGCTCTCAGCTCTGCTCAGTCACTCTTTTGGTTCACACA 840
 Db |||||
 Qy 906 AGGTTACGTGTGTACAACTCAGCTCTCAGCTCTGCTCAGTCACTCTTTTGGTTCACACA 965
 Db |||||
 Qy 841 TCCCAATTTAAAGATCTTGACAGCCTCCATACAGAAATCATGTGTCAAGCTTCAGTGATTG 900
 Db |||||
 Qy 966 TCCCAATTTAAAGATCTTGACAGCCTCCATACAGAAATCATGTGTCAAGCTTCAGTGATTG 1025
 Db |||||

QY	901	TGCTTATTACACAGCTATTCTTTGCTGCAACCTGATTGAAATGTTTCAGAGTTAGGCTT	960
Db	1026	TGCTTATTACACAGCTATTCTTTGCTGCAACCTGATTGAAATGTTTCAGAGTTAGGCTT	1085
QY	961	GACACCCACACATGTCATTTAGATTAATAAATAACATCTTTTGGAGTCTTTGGCATCAGTTATCT	1020
Db	1086	GACACCCACACATGTCATTTAGATTAATAAATAACATCTTTTGGAGTCTTTGGCATCAGTTATCT	1145
QY	1021	CAAGTCGTATACCTGTGGTTGTCGAATCCAGTGGACTTTTGTAGTACTCAGATCCCTCT	1080
Db	1146	CAAGTCGTAGACCTGTGGTTGTCGAATCCAGTGGACTTTTGTAGTACTCAGATCCCTCT	1205
QY	1081	TTGCTTTGGTCGTAGTATATTAACAAGTAAACCTGCTGTATGCTCACCAGAAAGAAAC	1140
Db	1206	TTGCTTTGGTCGTAGTATATTAACAAGTAAACCTGCTGTATGCTCACCAGAAAGAAAC	1265
QY	1141	AGAGCATGTCAAGTCTTTGATCTCTGAGAGCTTTTGTCTTACAAATTTCCAGAAATGCTGA	1200
Db	1266	AGAGCATGTCAAGTCTTTGATCTCTGAGAGCTTTTGTCTTACAAATTTCCAGAAATGCTGA	1325
QY	1201	AGGCCACATCAVTTTAGGAAAGTTTAATTGTACCTTTAAAGGTCAAAGAGTTATAGA	1260
Db	1326	AGGCCACATCAVTTTAGGAAAGTTTAATTGTACCTTTAAAGGTCAAAGAGTTATAGA	1385
QY	1261	TTCCACTGTATTACCTGGGATACCTATTGAAATGTGAGAGTTCAATTAATGAGGCTATT	1320
Db	1386	TTCCACTGTATTACCTGGGATACCTATTGAAATGTGAGAGTTCAATTAATGAGGCTATT	1445
QY	1321	ACCTATCAAAAAATCAACTGCCCTCAAGGTGGGACCTCTTTGTACAACTTTATCCGGAGA	1380
Db	1446	ACCTATCAAAAAATCAACTGCCCTCAAGGTGGGACCTCTTTGTACAACTTTATCCGGAGA	1505
QY	1381	CACCTCTGACACTGGAGAAGAACTGTGTGTGTGAGTTATGAGGTTTCTCTTGAAATGTC	1440
Db	1506	CACCTCTGACACTGGAGAAGAACTGTGTGTGTGAGTTATGAGGTTTCTCTTGAAATGTC	1565
QY	1441	AGTCTTGGACCGCTCTTAACCTAGGAAGGCGAGCTTAATCAGTGACGAGTACGTTCTGT	1500
Db	1566	AGTCTTGGACCGCTCTTAACCTAGGAAGGCGAGCTTAATCAGTGACGAGTACGTTCTGT	1625
QY	1501	CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGATCATCGTATTAT	1560
Db	1626	CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGATCATCGTATTAT	1685
QY	1561	TGCCATAGACAGAAATTGGAGTGACTCTGTATGGAAACCCCTGACTAAATGACAGGAACACA	1620
Db	1686	TGCCATAGACAGAAATTGGAGTGACTCTGTATGGAAACCCCTGACTAAATGACAGGAACACA	1745
QY	1621	GCCTATTGGATCCTAGGCTCAATATGCTCTAATAGTTATGGAAGTGTGAAAGATGTGTG	1680
Db	1746	GCCTATTGGATCCTAGGCTCAATATGCTCTAATAGTTATGGAAGTGTGAAAGATGTGTG	1805
QY	1681	CACGTCAAAATTTGGCTCCCAACATTTTTTTCATCTTATTCTTAATGAAGCAACAATCTG	1740
Db	1806	CACGTCAAAATTTGGCTCCCAACATTTTTTTCATCTTATTCTTAATGAAGCAACAATCTG	1865
QY	1741	CAGCTTGCTTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCACGTGTCA	1800
Db	1866	CAGCTTGCTTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCACGTGTCA	1925
QY	1801	GACGGCACTGCATGCTCTCGAGTTAACACTCAAGGAACCAATGGGCTTTGTTGGAGGTGG	1860
Db	1926	GACGGCACTGCATGCTCTCGAGTTAACACTCAAGGAACCAATGGGCTTTGTTGGAGGTGG	1985
QY	1861	CTGTACTGAAACTCATTTGGCTGCATATATCAGACACAAGACTCACACGCCAGAAAG	1920
Db	1986	CTGTACTGAAACTCATTTGGCTGCATATATCAGACACAAGACTCACACGCCAGAAAG	2045
QY	1921	CATTCTCAAAGATGATGAATGTACTCAAAACAGAACTTCAATTAATTTGCTGAAGCATTTTG	1980
Db	2046	CATTCTCAAAGATGATGAATGTACTCAAAACAGAACTTCAATTAATTTGCTGAAGCATTTTG	2105
QY	1981	CAGTCCCTTAGAATCTGTTTGGCTCTTTTAGAACATGATGAGGTGAAATTTCTCACTGA	2040
Db	2106	CAGTCCCTTAGAATCTGTTTGGCTCTTTTAGAACATGATGAGGTGAAATTTCTCACTGA	2165
QY	2041	CATGAAGTATGACACACCTTTTGGTCAGTTTCAGGCAGATTCTCCCTGTGTGTAACTGGCC	2100
Db	2166	CATGAAGTATGACACACCTTTTGGTCAGTTTCAGGCAGATTCTCCCTGTGTGTAACTGGCC	2225
QY	2101	AGATTTGCTTTTACAGTGTGGCTGTGGATTATACAAATAGCCAGGAAGAACTCAACTGGTC	2160
Db	2226	AGATTTGCTTTTACAGTGTGGCTGTGGATTATACAAATAGCCAGGAAGAACTCAACTGGTC	2285
QY	2161	TTTCTTAAAGAAGCACAGTCGTCCTTGTGTCACAAAGCTGCTTCCACATGAAGCTGT	2220
Db	2286	TTTCTTAAAGAAGCACAGTCGTCCTTGTGTCACAAAGCTGCTTCCACATGAAGCTGT	2345
QY	2221	GGGCTCAGCCAGCAACCTTGACCTTGGACTTGTGTGCTGCAAAAGCTTAGTGGCTACAGGT	2280
Db	2346	GGGCTCAGCCAGCAACCTTGACCTTGGACTTGTGTGCTGCAAAAGCTTAGTGGCTACAGGT	2405
QY	2281	GGCTGTAGAGACAGCCAAATTTGATTTGGGATCTTTTCATATGTTATTTGAAGATAAAACTA	2340
Db	2406	GGCTGTAGAGACAGCCAAATTTGATTTGGGATCTTTTCATATGTTATTTGAAGATAAAACTA	2465
QY	2341	ACAGAAATAGCATGTTGCTATTACAAAGAGAAACAAATAACTAGTCTGTTGGCAATTGAGA	2400
Db	2466	ACAGAAATAGCATGTTGCTATTACAAAGAGAAACAAATAACTAGTCTGTTGGCAATTGAGA	2525
QY	2401	AAAAAAAAAAAAA 2413	
Db	2526	AAAAAAAAAAAAA 2538	
RESULT 4			
AAA08567			
ID	AAA08567 standard; DNA; 2776 BP.		
XX	AC AAA08567;		
XX	DT 19-JUL-2000 (first entry)		
XX	Human chaperone protein 5 (HCHP-5) coding sequence.		
XX	Human chaperone protein; HCHP-5; neurodegenerative; metabolic;		
XX	developmental; autoimmune; inflammatory; cell proliferation; cancer; ss.		
XX	Homo sapiens.		
XX	Key Location/Qualifiers		
XX	CDS 748..2460		
XX	FT /*tag= a		
XX	FT /product= "HCHP-5"		
XX	WO200017358-A2.		
XX	30-MAR-2000.		
XX	22-SEP-1999; 99WO-US022027.		
XX	22-SEP-1998; 98US-00158642.		
XX	22-SEP-1998; 98US-0172221P.		
XX	19-JAN-1999; 99US-00233291.		
XX	19-APR-1999; 99US-00294698.		
XX	19-APR-1999; 99US-0172232P.		
XX	(INCV-) INCYTE PHARM INC.		
XX	Tang YT, Hillman JL, Yue H, Patterson C, Baughn MR, Batra S;		
XX	WPI; 2000-283583/24.		
XX	P-PSDB; AAY91944.		
XX	New purified polypeptides and polynucleotides encoding human chaperone		

PT	proteins, useful for diagnosing, treating and preventing disorders
PT	associated with the expression human chaperone proteins.
XX	
PS	Claim 7; Page 79-80; 88pp; English.
XX	
CC	AAA08563-68 encode human chaperone proteins 1-6 (HCHP-1 to HCHP-6)
CC	respectively. The sequences can be used to treat and prevent disorders
CC	associated with altered expression or activity of HCHP comprising
CC	administering a composition comprising the polypeptide or an antagonist
CC	to a patient (claimed). The human chaperone proteins are also useful for
CC	the diagnosis, treatment or prevention of neurodegenerative, metabolic,
CC	developmental, autoimmune/inflammatory disorders and cell proliferative
CC	disorders including cancer
XX	
SQ	Sequence 2776 BP; 834 A; 535 C; 593 G; 814 T; 0 U; 0 Other;
Query Match	97.1%; Score 2399.2; DB 3; Length 2776;
Best Local Similarity	99.9%; Pred. No. 0;
Matches 2401; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
Qy	1 TATGAGCCTTCGGAACTTGTGGAGAGACTACAAAGTTTGGTTGTTATGGTCCCTTTAGT 60
Db	120 TATGAGCCTTCGGAACTTGTGGAGAGACTACAAAGTTTGGTTGTTATGGTCCCTTTAGT 179
Qy	61 TGGGCTCATACATTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTTTTCCAAATACCTAA 120
Db	180 TGGGCTCATACATTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTTTTCCAAATACCTAA 239
Qy	121 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACCTTCAAAATCTTCAGAAAGCCAAAT 180
Db	240 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACCTTCAAAATCTTCAGAAAGCCAAAT 299
Qy	181 CCAGGGGAAGTAGCAGGCTTGGCAATCTTCAGGTAAAGAGCAGCTTTGAATCTGAGCTTC 240
Db	300 CCAGGGGAAGTAGCAGGCTTGGCAATCTTCAGGTAAAGAGCAGCTTTGAATCTGAGCTTC 359
Qy	241 ATATCGAAAGAGAGATGAAATAATACAGTGGATTAGAAAGAACTGGCTTCTTTGAGCT 300
Db	360 ATATCGAAAGAGAGATGAAATAATACAGTGGATTAGAAAGAACTGGCTTCTTTGAGCT 419
Qy	301 GGGATATCTTTCATAGTGTCCATCTTGGAAACATATCTTTTGAGAGGCTCGAAAGCAG 360
Db	420 GGGATATCTTTCATAGTGTCCATCTTGGAAACATATCTTTTGAGAGGCTCGCAAGCAG 479
Qy	361 TCTGTAATAATTTTCAGTCTCAAGCAACAAAGAGATTTGAAGAGTGAAGTAAATAAAT 420
Db	480 TCTGTAATAATTTTCAGTCTCAAGCAACAAAGAGATTTGAAGAGTGAAGTAAATAAAT 539
Qy	421 ATTTGGAAATTAATAATTTGTTCATTAATCAATCTATGCTGATTAGCTTCATTAACATTGA 480
Db	540 ATTTGGAAATTAATAATTTGTTCATTAATCAATCTATGCTGATTAGCTTCATTAACATTGA 599
Qy	481 ACTTTTGTATTTATAGCCCAATGCTGCATATTCATCTTAATTTCTTAAGAAATAAT 540
Db	600 ACTTTTGTATTTATAGCCCAATGCTGCATATTCATCTTAATTTCTTAAGAAATAAT 659
Qy	541 TTTAAATGTTAAACGTGATAATGCAATAAATAGAAAATGTTGTTTACAAAATAAAGACG 600
Db	660 TTTAAATGTTAAACGTGATAATGCAATAAATAGAAAATGTTGTTTACAAAATAAAGACG 719
Qy	601 GTCTTCACATAGTTACCACTGAAGATAGTCTCGTTTGGAAAGCTGAAGAGCCATCAT 660
Db	720 GTCTTCACATAGTTACCACTGAAGATAGTCTCGTTTGGAAAGCTGAAGAGCCATCAT 779
Qy	661 GTGTAGAGTGAACCACTGACCACTGAGAGAGTCAGGACCACTTTCTGTCTTGAAG 720
Db	780 GTGTAGAGTGAACCACTGACCACTGAGAGAGTCAGGACCACTTTCTGTCTTGAAG 839
Qy	721 AATTGTAACATCATGCTATGSCCCCTCAGGTAGGCTGAAGCAGCTGCACAATGGCTTTG 780
Db	840 AATTGTAACATCATGCTATGSCCCCTCAGGTAGGCTGAAGCAGCTGCACAATGGCTTTG 899
Qy	781 AGGTTACGTGTGTACAACTTCACAGTCTCAGCTCTGCTCAGTCACCTTTTGGTCAACA 840

Db	900 AGTTACGTGTGTACAACTTCACAGTCTCTGCTCAGTCACCTTTTGGTCAACA 959
Qy	841 TCCCATTTTAAAGATCCCTGACAGCCTCCATACAGATCATGTGTCAAGCTTCAGTCAT 900
Db	960 TCCCATTTTAAAGATCCCTGACAGCCTCCATACAGAAATCATGTGTCAAGCTTCAGTCAT 1019
Qy	901 TGGCTTATTTACAGAGCTATTTCTTTGTGCAACCTGATTGAAAATGTTTCAGAGATTAGGCT 960
Db	1020 TGGCTTATTTACAGAGCTATTTCTTTGTGCAACCTGATTGAAAATGTTTCAGAGATTAGGCT 1079
Qy	961 GACACCCACATCTGTCATTTAGATTAAATAAAACATCTTTTGAGTCTTTTGATCATGTTATCT 1020
Db	1080 GACACCCACATCTGTCATTTAGATTAAATAAAACATCTTTTGAGTCTTTTGATCATGTTATCT 1139
Qy	1021 CAACTCTGATACCTGCTGTTGCGAATCCAGTGGACCTTTAGTAGTACTCAGATCCCTCT 1080
Db	1140 CAACTCTGAGACTGCTGGTTGCGAATCCAGTGGACCTTTAGTAGTACTCAGATCCCTCT 1199
Qy	1081 TTTGTTGGTGGCTAGTATATTAAACAAAGTAAACCTGCTCTATGCTCACCAAGAAAGAAAC 1140
Db	1200 TTTGTTGGTGGCTAGTATATTAAACAAAGTAAACCTGCTCTATGCTCACCAAGAAAGAAAC 1259
Qy	1141 AGAGCATGTCAAGTCTTTGATCCTCAGAGCCTTTTGTCTTACAATTCAGAAAATGCTGA 1200
Db	1260 AGAGCATGTCAAGTCTTTGATCCTCAGAGCCTTTTGTCTTACAATTCAGAAAATGCTGA 1319
Qy	1201 AGGCCACATCATTTTAGGAAAGAGTTAATTTGCTTCTTAAAGGTCAAAGAGTTATAGA 1260
Db	1320 AGGCCACATCATTTTAGGAAAGAGTTAATTTGCTTCTTAAAGGTCAAAGAGTTATAGA 1379
Qy	1261 TTTCCACTGTATTAACCTGGGATACCTCAATTGAAATGTGAGAAATCAATTAATGAGGCTATT 1320
Db	1380 TTTCCACTGTATTAACCTGGGATACCTCAATTGAAATGTGAGAAATCAATTAATGAGGCTATT 1439
Qy	1321 ACCTATCAAAAAATCAATGCCCCCTCAAGGTGGCACTCTTTTGTACAACCTTTATCCGGAGA 1380
Db	1440 ACCTATCAAAAAATCAATGCCCCCTCAAGGTGGCACTCTTTTGTACAACCTTTATCCGGAGA 1499
Qy	1381 CACTTCTGACCTGGAGAGGAACTGTGGTGGTCAAGTATGGGTTCTCTTGGAAATGTC 1440
Db	1500 CACTTCTGACCTGGAGAGGAACTGTGGTGGTCAAGTATGGGTTCTCTTGGAAATGTC 1559
Qy	1441 AGCTTTGGACCAAGCTGCTTTAACTTAGGAAGGAGCTTAATCAGTGACCACTAGATCTTGT 1500
Db	1560 AGCTTTGGACCAAGCTGCTTTAACTTAGGAAGGAGCTTAATCAGTGACCACTAGATCTTGT 1619
Qy	1501 CCGTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGCTATATTAT 1560
Db	1620 CCGTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGCTATATTAT 1679
Qy	1561 TGCCATAGACAGAAATGGAGTGACTCTGATGGAAACCCCTGACTTAATGACAGGACACA 1620
Db	1680 TGCCATAGACAGAAATGGAGTGACTCTGATGGAAACCCCTGACTTAATGACAGGACACA 1739
Qy	1621 GCCTATTGGATCCCTAGGCTCAATATGCTCAATAGTTATGGAAGTGTGAAAGATGTGTG 1680
Db	1740 GCCTATTGGATCCCTAGGCTCAATATGCTCAATAGTTATGGAAGTGTGAAAGATGTGTG 1799
Qy	1681 CACTGCAAAATTTGGCTCCCAACATTTTCTTCTTAATGAAACCAATCTG 1740
Db	1800 CACTGCAAAATTTGGCTCCCAACATTTTCTTCTTAATGAAACCAATCTG 1859
Qy	1741 CAGCTTGTCTTCTGCAACAGAAATGACATGCTCCCTGGGATGAGCTCAAGCTCAAGTGTCA 1800
Db	1860 CAGCTTGTCTTCTGCAACAGAAATGACATGCTCCCTGGGATGAGCTCAAGCTCAAGTGTCA 1919
Qy	1801 GACGCACTGCAATGCTTGGCAGTTAACTCAAGGAACCTAGGCTTTGTTGGAGGCTG 1860
Db	1920 GACGCACTGCAATGCTTGGCAGTTAACTCAAGGAACCTAGGCTTTGTTGGAGGCTG 1979
Qy	1861 CTGTACTGAAACTCAATTTGGCTGCTATATACAGACAAAGACTCAACACCCAGAAAG 1920

Db 1980 CTGACTGAACTCATTTGGCTGCATATATACAGACACAGACTCACACGCCAGAAAG 2039
QY 1921 CATTCTCAAAGATGATGAATGTAATCTCAACAGAACTTCAATTAATTGCTGAAGCAATTTG 1980
Db 2040 CATTCTCAAAGATGATGAATGTAATCTCAACAGAACTTCAATTAATTGCTGAAGCAATTTG 2099
QY 1981 CAGTGGCCCTAGATCTGTTGGCTCTTTAGACATGATGGAGGTGAATTTCTACTGA 2040
Db 2100 CAGTGGCCCTAGATCTGTTGGCTCTTTAGACATGATGGAGGTGAATTTCTACTGA 2159
QY 2041 CATGAAGTAGGACACCTTTGGTTCAGTTCAGGACAGATTTCCCTGTGTGCTAACTGGCC 2100
Db 2160 CATGAAGTAGGACACCTTTGGTTCAGTTCAGGACAGATTTCCCTGTGTGCTAACTGGCC 2219
QY 2101 AGATTTGCTTTACAGTGGCTGTGTGGATPATACAATAGCCAGGAAGAACTCAACTGGTC 2160
Db 2220 AGATTTGCTTTACAGTGGCTGTGTGGATPATACAATAGCCAGGAAGAACTCAACTGGTC 2279
QY 2161 TTTCTTAAGNAGCACAGCTGCTCCATTTGTCACCAAGCTGCTTCCACATGAAGCTGT 2220
Db 2280 TTTCTTAAGNAGCACAGCTGCTCCATTTGTCACCAAGCTGCTTCCACATGAAGCTGT 2339
QY 2221 GGGCTCAGCCAGCAACCTGACCTTGGACTGTTTGAATGCTGCAAGCTTAGTGGCTACAGGT 2280
Db 2340 GGGCTCAGCCAGCAACCTGACCTTGGACTGTTTGAATGCTGCAAGCTTAGTGGCTACAGGT 2399
QY 2281 GGGCTCAGCCAGCAACCTGACCTTGGACTGTTTGAATGCTGCAAGCTTAGTGGCTACAGGT 2340
Db 2400 GGGCTCAGCCAGCAACCTGACCTTGGACTGTTTGAATGCTGCAAGCTTAGTGGCTACAGGT 2459
QY 2341 AGAGAATAGCATCTGCTATTACAGAGAAACAATTAATTAATGTTTGAAGATATAAACTA 2400
Db 2460 AGAGAATAGCATCTGCTATTACAGAGAAACAATTAATTAATGTTTGAAGATATAAACTA 2519
QY 2401 AAAAA 2404
Db 2520 AAAAA 2523

RESULT 5
AAS83823
ID AAS83823 standard; cdna; 2562 BP.
XX
AC AAS83823;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #19627.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Dmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG19636.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess

PT biodiversity.
XX Claim 1; SEQ ID NO 19627; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (II) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have application in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2562 BP; 755 A; 510 C; 561 G; 736 T; 0 U; 0 Other;

Query Match 79.4%; Score 1961.6; DB 5; Length 2562;
Best Local Similarity 95.3%; Pred. No. 2.8e-314;
Matches 2340; Conservative 0; Mismatches 69; Indels 47; Gaps 29;

QY 1 TATGAGCCTTCGGAACCTTGTGAGAGACTACAAA--GTTTTGGTGTATGGTCCCTTTA 58
Db 107 TATGAGCCTTCGGAACCTTGTGAGAGACTACAAAAGTTTGGGTGTATGGTCCCTTTA 166
QY 59 GTTGGGCTCATACATTTGGGGTGGTACAGAAATCAAGCA--GCCCTGTTTCCCAATACC 117
Db 167 GTTGGGCTCATACATTTGGGGTGGTACAGAAATCAAGCAAGCGCCTTGTTTCCCAATACC 226
QY 118 TAAAAACGACGACATTCCTTGAGCAAGA--TAGTCTGGGACTTT--CAAAATCTTCAGAAGAG 174
Db 227 TAAAAACGACGACATTCCTTGAGCAAGAATAGTCTGGGACTTTTCAATCTTCAGAAGAG 286
QY 175 CC-AAATCCAGGGGAAGTAGCAGGCTTGCAATCTTCAGGT-AAAGAAGCAGC-TTTGAAT 231
Db 287 CCAAAATCCAGGGGAAGTAGCAGGCTTGCAATCTTCAGGTAAAGAAGCAGCTTTGAAT 346
QY 232 CTGAGCTTCATAT-CGAAAGAGAGATGAAGAA--TACCAGTTGGATTA--GAAAGAACT 286
Db 347 CTGAGCTTCATATCCGAAAGAGAGATGAAGAAATACCAAGTTGGATTAAGAAGAACTT 406
QY 287 GGCCTCTCTAGCTGGGATATC---TTTCATAGGTGTCCATC-TTGGAAACATCTTTTTT 341
Db 407 GGCCTCTCTAGCTGGGATATCCTTTCCATAGGTGTCCATCTTTGGAACATCTTTTT 466
QY 342 GCAGAGGTC-TGCAAGCAGTC-TGTAAATTTTCAAGTTCAGCAAGCAACAAAGAG--T 396
Db 467 GCAGAGGTC-TGCAAGCAGTC-TGTAAATTTTCAAGTTCAGCTCCAAAGCAACAAAGAGTA 526
QY 397 ATTGAAGAGTGAAGTAAATTAATAT-TTGGAAATTAATTTTGTATAGCCCAATTCATTTCA 455
Db 527 TTGAAGAGTGAAGTAAATTAATATTTTGGGATTAATTTTGTATAGCCCAATTCATTTCA 586
QY 456 TGCTGATTAGCTTCATAAACATTTTGAATTTTATAGCCCAATTCGTCATTTTC 515
Db 587 TGCTGATTAGCTTCATAAACATTTTGAATTTTATAGCCCAATTCGTCATTTTC 646
QY 516 ATACTTTAATTCCTAAAGAAATTAATTTTAAATGTTAAACGTCATTAATGAATAATAGAA 575
Db 647 ATACTTTAATTCCTAAAGAAATTAATTTTAAATGTTAAACGTCATTAATGAATAATAGAA 706
QY 576 AAATGTGGTTTACAAAATAAAAAACGCTCTTCACTAGTTTACCACCTGGAAGTAAAGTGTCTC 635

Db 707 AAATGTGGTTTACAAAATAAAGCGGTCTTCACTAGTTACCACTGAAGTAAGATGTCTC 766
Qy 636 GTTTGGAAGCTAAGAAGCCATCATTTGTGTAAGAGTGAACCACTGACAACTGAGAGAGTCA 695
Db 767 GTTTGGAAGCTAAGAAGCCATCATTTGTGTAAGAGTGAACCACTGACAACTGAGAGAGTCA 826
Qy 696 GGACCAACATTTCTGTCTTGAAGAAGAAATGTAACATCATGCTATGAGGCCCCCTCAGGTAGGC 755
Db 827 GGACCAACATTTCTGTCTTGAAGAAGAAATGTAACATCATGCTATGAGGCCCCCTCAGGTAGGC 886
Qy 756 TGAAGCAGCTGCACAAATGGCTTTGGAGGTTACGTGTGTACAACTCAGAGTCTCAGCTC 815
Db 887 TGAAGCAGCTGCACAAATGGCTTTGGAGGTTACGTGTGTACAACTCAGAGTCTCAGCTC 946
Qy 816 TGCTCAGTCACTTTTGGTGCACATCCCATTTTAAAGATCTTGAAGATCTTGAAGATCTTGAAG 875
Db 947 TGCTCAGTCACTTTTGGTGCACATCCCATTTTAAAGATCTTGAAGATCTTGAAGATCTTGAAG 1006
Qy 876 ATCATGTGTCAAGCTTCAGTGAATGTGGCTTATTCACAGCTATTTCTTGTGCTCAACCTGA 935
Db 1007 ATCATGTGTCAAGCTTCAGTGAATGTGGCTTATTCACAGCTATTTCTTGTGCTCAACCTGA 1066
Qy 936 TTGAAAATGTTACAGAGATAGGCTTGACACCCACCACTGTCATTTAGATTAATAAATCAATC 995
Db 1067 TTGAAAATGTTACAGAGATAGGCTTGACACCCACCACTGTCATTTAGATTAATAAATCAATC 1126
Qy 996 TTTTGTAGTCTTGCATCAGTATATCTCAAGTCTCATACCTGTGTGTCGAACTCCAGTGG 1055
Db 1127 TTTTGTAGTCTTGCATCAGTATATCTCAAGTCTCATACCTGTGTGTCGAACTCCAGTGG 1186
Qy 1056 ACTTTAGTAGTACTCAGATCCTCTCTTTGTTGGTGGGTAGTATATTAACAAGTAAACCTG 1115
Db 1187 ACTTTAGTAGTACTCAGATCCTCTCTTTGTTGGTGGGTAGTATATTAACAAGTAAACCTG 1246
Qy 1116 CCTGTATGCTCACAGAAAGGAAACAGAGATGTCTAGTGTCTTGTGATCTTGTGAGAGCTTTT 1175
Db 1247 CCTGTATGCTCACAGAAAGGAAACAGAGATGTCTAGTGTCTTGTGATCTTGTGAGAGCTTTT 1306
Qy 1176 TGCTTCAATTTCCAGAAAATGCTGAAGGCCACATCAATTTTAGGAAAGAGTTTAAATGTATC 1235
Db 1307 TGCTTCAATTTCCAGAAAATGCTGAAGGCCACATCAATTTTAGGAAAGAGTTTAAATGTATC 1366
Qy 1236 CTTTAAAGGTCAAGAGTTATAGATTCCACTGTATTAACCTGGGATCTCAATTTGAATGT 1295
Db 1367 CTTTAAAGGTCAAGAGTTATAGATTCCACTGTATTAACCTGGGATCTCAATTTGAATGT 1426
Qy 1296 CAGAAGTTCAATTAATGAGCTATTACCTATCAAAAAATCAACTGCCCTCAAGGTGGCAC 1355
Db 1427 CAGAAGTTCAATTAATGAGCTATTACCTATCAAAAAATCAACTGCCCTCAAGGTGGCAC 1486
Qy 1356 TCTTTTGTACAACCTTTATCCGGAGACATTC-TGACACTGGAGAGAACTGTGGTGTGTC 1414
Db 1487 TCTTTTGTACAACCTTTATCCGGAGACATTCGTGTGACACTGGAGAGAACTGTGGTGTGTC 1546
Qy 1415 AGTTATGGGGTTTCTCTTGAATAATGCAGTCTTGGACAGCTGTAACTAGAGAGGACAG 1474
Db 1547 AGTTATGGGGTTTCTCTTGAATAATGCAGTCTTGGACAGCTGTAACTAGAGAGGACAG 1606
Qy 1475 CTAATCAGTGACACCGTAGATCTTGTCTGTGCCAAAAAGTTATACATCCATCTTTTGAAG 1534
Db 1607 CTAATCAGTGACACCGTAGATCTTGTCTGTGCCAAAAAGTTATACATCCATCTTTTGAAG 1666
Qy 1535 CAGTTTCTCAATATGAT-CGTATTATTCAGATGACAGAAATTT-GGAGTACCTCTGATGG 1592
Db 1667 CAGTTTCTCAATATGATCGTATTATTCATAGACAGAAATTTGGGAGTACCTCTGATGG 1726
Qy 1593 AACCCC----TGACTAAAATGACAGGAAACAGACCTTATTTGGATCCCTAGGCTCAATATGT 1648
Db 1727 GAACCCCTGGACTTAAATGACAGGAAACAGACCTTATTTGGATCCCTAGGCTCAATATGT 1786
Qy 1649 CCTAATAGTTATGGAAATGTGAAGATGTGTGCATCTGCAAAAATTTGGCTCCCAACATTTT 1708

Db 1787 CCTAATAGTTATGGAAGTGTGAAAGATGTGTGCATCTGCAAAAATTTGGCTCCAAACATTTT 1846
Qy 1709 TTTTCATCTTTATTCCTAATGAAGCAAAATCTCAGCTTCTCTCTCTGCAACAGAAATGAC 1768
Db 1847 TTTTCATCTTTATTCCTAATGAAGCAAAATCTCAGCTTCTCTCTCTGCAACAGAAATGAC 1906
Qy 1769 ACTGCTGGGATGAGCTGAAAGCTCAGCTGTGAGAGCGCACTGCAATCTCTGCAAGTTAAACA 1828
Db 1907 ACTGCTGGGATGAGCTGAAAGCTCAGCTGTGAGAGCGCACTGCAATCTCTGCAAGTTAAACA 1966
Qy 1829 CTCAGGAACCAATGGGCTTTGTTGGAGGTGGCTGTACTGAAACATCATTTGGCTGCATAT 1888
Db 1967 CTCAGGAACCAATGGGCTTTGTTGGAGGTGGCTGTACTGAAACATCATTTGGCTGCATAT 2026
Qy 1889 ATCAGACACAAGACTCACAAACGACCCAGAAAGCAATTTCTCAAGATGATGAATGACTCTCAA 1948
Db 2027 ATCAGACACAAGACTCACAAACGACCCAGAAAGCAATTTCTCAAGATGATGAATGACTCTCAA 2086
Qy 1949 ACAGAACTTTCAATTAATTTGCTGAAGCAATTTTGCAGTGGCTTGAATCTCTGTTGGCTCT 2008
Db 2087 ACAGAACTTTCAATTAATTTGCTGAAGCAATTTTGCAGTGGCTTGAATCTCTGTTGGCTCT 2146
Qy 2009 TTAGAACATGATGGAGGTGAAATTTCTC-ACTGACATGAAGTATGGACACCTTTGGTCACT 2067
Db 2147 TTAGAACATGATGGAGGTGAAATTTCTCAACTGACATGAAGTATGGACACCTTTGGTCACT 2206
Qy 2068 TCAGGCAGATTTCTCCC---TGTTGTTGCTAACTGGCCAGATTTGCTTTCACAGTGTGGCTG 2124
Db 2207 TCAGGCAGATTTCTCCCCTGGGTTGCTGAACCTGGCCAGATTTGCTTTCACAGTGTGGCTG 2266
Qy 2125 TGGATTATACATAGCCAGAAAGAACTCAACTGGTCTTTCTTT-AAGAAGCACACCTG-CGT 2182
Db 2267 TGGATTATACATAGCCAGAAAGAACTCAACTGGTCTTTCTTTAAGAAGCACACCTGCCGT 2326
Qy 2183 CCATTTGTGCCCAAAAGCT--GCCTTCCACATGAAGTGTGGGCTCAGCCA-GCAACCTG 2239
Db 2327 CCATTTGTGCCCAAAAGCTTGGCTTTCCACATGAAGTGTGGGCTCAGCCAGCAACCTG 2386
Qy 2240 ACCTTTGGAC-TGTTTGTACTGCAAAAGCTT---AGTGGCTTACAGTGTGGCTGTAGACAGC 2295
Db 2387 ACCTTTGGACTTGTGGCTGCAAAAGCTTTAGTGGGCTTACAGGTGGCTGTAGACAGAC 2446
Qy 2296 CAATTTGATTTGGGATCTTTT-CATATGTTATTGAAGATAAAAACTAAGAGAAATAGCATG 2353
Db 2447 CAATTTGATTTGGGCTCTTTTCATATTTGTTATTGAAGATAAAAACTAAGAGAAATAGCATG 2506
Qy 2354 TTCTGATTACAGAGAAACAAATAAATAGTCTGTGGCAATTGAGAAAAAATAA 2409
Db 2507 TTCTGATTACAGAGAAACAAATAAATAGTCTGTGGCAATTGAGAAAAAATAA 2562

RESULT 6
AAC99959
ID AAC99959 standard; cdna; 1921 BP.
XX
AC AAC99959;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human secreted protein gene 142 SEQ ID NO:152.
XX
KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; gene therapy; pathological condition;
KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
KW cerebrovascular disorder; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; skin aging; food additive; preservative; ss.
OS
XX Homo sapiens.

PN WO200070042-A1.
XX 23-NOV-2000.
PD
XX 11-MAY-2000; 2000WO-US012788.
PF
XX 13-MAY-1999; 99US-0134068P.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
PI Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;
PI Duan RD, Florence KA, Soppet DR;
XX
XX WPI: 2000-679828/66.
XX P-PSDB; AAB56218.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is used
XX in preventing, treating or ameliorating a medical condition.
XX
XX Claim 1; Page 930-931; 1065pp; English.
XX
XX The polynucleotide sequences given in AAC99818 to AAC99977 encode the
XX human secreted proteins given in AAB56077 to AAB56362. Human secreted
XX proteins have activities based on the tissues and cells the genes are
XX expressed in. Examples of activities include: immunosuppressive;
XX antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
XX vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
XX virucide; fungicide; and ophthalmological. The human secreted
XX polynucleotides and proteins can be used to prevent, treat or ameliorate
XX a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
XX dogs, chickens or sheep. They are also used in diagnosing a pathological
XX condition or susceptibility to a pathological condition. Disorders which
XX are diagnosed or treated include autoimmune diseases e.g. rheumatoid
XX arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
XX liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
XX disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
XX e.g. Alzheimer's disease, infections caused by bacteria, viruses and
XX fungi and ocular disorders e.g. corneal infection. The proteins can also
XX be used to aid wound healing and epithelial cell proliferation, to
XX prevent skin aging due to sunburn, to maintain organs before
XX transplantation, for supporting cell culture of primary tissues, to
XX regenerate tissues and in chemotaxis. The proteins can also be used as a
XX food additive or preservative to increase or decrease storage
XX capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
XX in the exemplification of the present invention
XX
SQ Sequence 1921 BP; 565 A; 380 C; 414 G; 556 T; 0 U; 6 Other;

Query Match 71.3%; Score 1763.2; DB 3; Length 1921;
Best Local Similarity 99.6%; Pred. No. 1.6e-281;
Matches 1765; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 TATGAGCCTTCGGAACCTTGTGGAGAGACTACAAAGTTTGGTTGTTATGGTCCCTTAGT 60
DB |||||
DB 126 TATGAGCCTTCGGAACCTTGTGGAGAGACTACAAAGTTTGGTTGTTATGGTCCCTTAGT 185
QY 61 TGGGCTCATACATTTGGGGTGTACAGAAATCAAGAGCCCTGTTTCCAAATACCTAA 120
DB |||||
DB 186 TGGGCTCATACATTTGGGGTGTACAGAAATCAAGAGCCCTGTTTCCAAATACCTAA 245
QY 121 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTCAAATCTTCAGAAAGCCAAAT 180
DB |||||
DB 246 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTCAAATCTTCAGAAAGCCAAAT 305
QY 181 CCAGGGGAAGTAGCAGCTTGCATCTTCAGGTAAAGAGCAGCTTTGAAATCTGAGCTTC 240
DB |||||
DB 306 CCAGGGGAAGTAGCAGCTTGCATCTTCAGGTAAAGAGCAGCTTTGAAATCTGAGCTTC 365
QY 241 ATATCGAAAGACAGATGAATAATACAGTTGAAATAGAAAGACCTGGCTTCTTTAGCT 300
DB |||||
DB 366 ATATCGAAAGACAGATGAATAATACAGTTGAAATAGAAAGACCTGGCTTCTTTAGCT 425
QY 301 GGGATATCTTTCATAGGTGTCATCTTGGAAACATACCTTTTTCGAGAGGTCTGCAAGCAG 360

DB |||||
DB 426 GGGATATCTTTCATAGGTGTCATCTTGGAAACATACCTTTTTCGAGAGGTCTGCAAGCAG 485
QY |||||
QY 361 TCTGTAATAATTTTCAGTCTCAAAAGCAAAAGAGATTTGAAGAGTGAAAGTAAAAATAAT 420
DB |||||
DB 486 TCTGTAATAATTTTCAGTCTCAAAAGCAAAAGAGATTTGAAGAGTGAAAGTAAAAATAAT 545
QY |||||
QY 421 ATTGGAATTTACTTAATTTGTCAATTAATCATTTCTATGCTGATTTAGCTTCATAACATTGA 480
DB |||||
DB 546 ATTGGAATTTACTTAATTTGTCAATTAATCATTTCTATGCTGATTTAGCTTCATAACATTGA 605
QY |||||
QY 481 ACTTTTGTATTTTATAGCCACAAATGCTGCATATTCATACCTTTAAATTCCTTAAAGAAATAAT 540
DB |||||
DB 606 ACTTTTGTATTTTATAGCCACAAATGCTGCATATTCATACCTTTAAATTCCTTAAAGAAATAAT 565
QY |||||
QY 541 TTTAATGTTTAAACGCTGATAATGCAATAAATAGAAAAATGTGGTTTACAAAAATAAAACG 600
DB |||||
DB 666 TTTAATGTTTAAACGCTGATAATGCAATAAATAGAAAAATGTGGTTTACAAAAATAAAACG 725
QY |||||
QY 601 GTCTTCACTAGTTACCACTGAAAGTAAAGTGTCTGTTTGGAGCTTAAAGAGCCATCATTT 560
DB |||||
DB 726 GTCTTCACTAGTTACCACTGAAAGTAAAGTGTCTGTTTGGAGCTTAAAGAGCCATCATTT 785
QY |||||
QY 661 GTGTAAGAGTGAAACCACTGACAACTGAGAGAGTTCAGGACCAACATTTCTGTCTTGAAGAG 720
DB |||||
DB 786 GTGTAAGAGTGAAACCACTGACAACTGAGAGAGTTCAGGACCAACATTTCTGTCTTGAAGAG 845
QY |||||
QY 721 AATTGTAAACATCATGCTATGCGCCCTCAGGTAGGTGAAGAGCTGCACAAATGGCTTTGG 780
DB |||||
DB 846 AATTGTAAACATCATGCTATGCGCCCTCAGGTAGGTGAAGAGCTGCACAAATGGCTTTGG 905
QY |||||
QY 781 AGGTTACGTTGTACAACTGACAGTCTCCTCAGCTCTGCTCAGTCACTTTTGGTTCACACA 840
DB |||||
DB 906 AGGTTACGTTGTACAACTGACAGTCTCCTCAGCTCTGCTCAGTCACTTTTGGTTCACACA 965
QY |||||
QY 841 TCCCATTTTAAAGATCTTCGACAGCCCTCCATACAGAAATCATGTGTCAAGCTTTCAGTGATTG 900
DB |||||
DB 966 TCCCATTTTAAAGATCTTCGACAGCCCTCCATACAGAAATCATGTGTCAAGCTTTCAGTGATTG 1025
QY |||||
QY 901 TGGCTTATTCACAGCTATTTCTTTGTCGAACTGAAATGTTGAAAATGTTTCAGAGATTAGGCTT 960
DB |||||
DB 1026 TGGCTTATTCACAGCTATTTCTTTGTCGAACTGAAATGTTGAAAATGTTTCAGAGATTAGGCTT 1085
QY |||||
QY 961 GACACCCACCACTGTCATTTAGATTAATAAACAATCTTTTGGCTCTTTCAGTCACTATCT 1020
DB |||||
DB 1086 GACACCCACCACTGTCATTTAGATTAATAAACAATCTTTTGGCTCTTTCAGTCACTATCT 1145
QY |||||
QY 1021 CAAAGTCTGATACCTGTGGTTGTGCAATCCAGTGGACTTTTAGTAGTACTCAGATCCTCCT 1080
DB |||||
DB 1146 CAAAGTCTGAGACCTGTGGTTGTGCAATCCAGTGGACTTTTAGTAGTACTCAGATCCTCCT 1205
QY |||||
QY 1081 TTGTTTGGTGGCTAGTATTAATAAAGTAAACCTGCTGTATGCTCAACAGAAAGAAAC 1140
DB |||||
DB 1206 TTGTTTGGTGGCTAGTATTAATAAAGTAAACCTGCTGTATGCTCAACAGAAAGAAAC 1265
QY |||||
QY 1141 AGAGCATGTGCTGCTTTGATCCTCAGAGCCCTTTTGGCTTACAATTCAGAAAAATGCTCA 1200
DB |||||
DB 1266 AGAGCATGTGCTGCTTTGATCCTCAGAGCCCTTTTGGCTTACAATTCAGAAAAATGCTCA 1325
QY |||||
QY 1201 AGGCCACATCATTTTAGGAAAGAGTTTAAATGTTGACCTTTTAAAGGTCAAAGAGTTATAGA 1260
DB |||||
DB 1326 AGGCCACATCATTTTAGGAAAGAGTTTAAATGTTGACCTTTTAAAGGTCAAAGAGTTATAGA 1385
QY |||||
QY 1261 TTCCACTGTATTTACCTGGGATATCTCATTTGAAATGTCAGAAAGTTCAATTAATGAGGCTATT 1320
DB |||||
DB 1386 TTCCACTGTATTTACCTGGGATATCTCATTTGAAATGTCAGAAAGTTCAATTAATGAGGCTATT 1445
QY |||||
QY 1321 ACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACCTTTATCCGGAGA 1380
DB |||||
DB 1446 ACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACCTTTATCCGGAGA 1505
QY |||||
QY 1381 CACTTCTGCACTGGAGAGGAACCTGTGTTGTTGCTGAGTTATGGGGTTTCTCTTGAATAATGC 1440
DB |||||

Db 1506 CACTTCGACATGGAGAGAACTGGTGGTCAGTTATGAGGGTTCTCTTCAAAATGC 1565
AAC77414
Qy 1441 AGCTTTGGACAGCTGCTTAACCTAGGAAGCGAGCTAATCAGTGACCACTAGATCTTGT 1500
Db 1566 AGCTTTGGACAGCTGCTTAACCTAGGAAGCGAGCTAATCAGTGACCACTAGATCTTGT 1625
Qy 1501 CCTGTGCAAAAAGTTATACATCATCTTTTGAAGCAGTTTCTCAATATGATCGTATTAT 1560
Db 1626 CCTGTGCAAAAAGTTATACATCATCTTTTGAAGCAGTTTCTCAATATGATCGTATTAT 1685
Qy 1561 TGCATAGACAGAAATGGAGTGACTCGATGGACCCCTGACTAATAATGACAGGACACA 1620
Db 1686 TGCATAGACAGAAATGGAGTGACTCGATGGACCCCTGACTAATAATGACAGGACACA 1745
Qy 1621 GCCTATTGGATCCCTAGGCTCAATATGCTCTTAATAGTTATGGAAGTGTGAAAGATGTGTG 1680
Db 1746 GCCTATTGGATCCCTAGGCTCAATATGCTCTTAATAGTTATGGAAGTGTGAAAGATGTGTG 1805
Qy 1681 CACTGCAAAATTTGGCTCCCAACATTTTTTTTCTATTCTTAATCCTAATGAACCAATCTG 1740
Db 1806 CACTGCAAAATTTGGCTCCCAACATTTTTTTTCTATTCTTAATCCTAATGAACCAATCTG 1865
Qy 1741 CAGCTTGCTCTCTGCAACAGAAATGACACTG 1772
Db 1866 CAGCTTGCTCTCTGCAACAGAAATGAGGTG 1897

RESULT 7

ID AAC77414 standard; cDNA; 1536 BP.

XX AAC77414;

DT 08-FEB-2001 (first entry)

XX Human ORFX ORP2969 polynucleotide sequence SEQ ID NO:5937.

KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoaric; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.

OS Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US008621.

XX 31-MAR-1999; 99US-0127607P.

XX 02-APR-1999; 99US-0127636P.

XX 05-APR-1999; 99US-0127728P.

XX 30-MAR-2000; 2000US-00540763.

XX (CURA-) CURAGEN CORP.

XX Shimketa RA, Leach M;

XX WPI; 2000-602362/57.

XX P-PSDB; AAB43205.

PT Novel nucleic acids and peptides derived from open reading frame x,

PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
PS
XX Claim 5; Page 5120-5121; 5507pp; English.

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipsoaric; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
XX Sequence 1536 BP; 457 A; 309 C; 324 G; 445 T; 0 U; 1 Other;

Query Match 60.8%; Score 1503.8; DB 3; Length 1536;
Best Local Similarity 99.2%; Pred. No. 9.1e-239;
Matches 1522; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

Qy 886 AAGCTTCAGTGATTTGGCTTATTCACAGCTATTCTTCTGCAACCTGATGAAATGT 945
Db 2 AAGCTTTAGTGATTTGGCTTATTCACAGCTATTCTTCTGCAACCTGATGAAATGT 61
Qy 946 TCAGAGATTAGGCTTGACACCCACCACCTGCTCATTAGATTAAATAAATCTTTTGAAGTCT 1005
Db 62 TCAGAGATTAGGCTTGACACCCACCACCTGCTCATTAGATTAAATAAATCTTTTGAAGTCT 121
Qy 1006 TTGCATCAGTTATCTCAA-GTCTGATACCTGTGGTTCGAATCCAGTGAGACTTTAGTA 1064
Db 122 TTGCATCAGTTATCTCAAAGGCTCGAGACCTGTGGTTCGAACCCAGTGAGACTTTAGTA 181
Qy 1065 GTACTCAGATCTCTCTTTCTTTGGTGGTAGTATATTAACAAGTAAACCTGCTGTATGC 1124
Db 182 GTACTCAGATCTCTCTTTCTTTGGTGGTAGTATATTAACAAGTAAACCTGCTGTATGC 241
Qy 1125 TCACAGAAAGGAAACAGAGCATGTCAGTGTCTTTGATCTCTGAGAGCTTTTGTCTTACAA 1184
Db 242 TCACAGAAAGGAAACAGAGCATGTCAGTGTCTTTGATCTCTGAGAGCTTTTGTCTTACAA 301
Qy 1185 TTCCAGAAATGCTGAAGGCCACATCATTTTGAAGAAAGTTTAAATGTACCTTTTAAAG 1244
Db 302 TTCCAGAAATGCTGAAGGCCACATCATTTTGAAGAAAGTTTAAATGTACCTTTTAAAG 361
Qy 1245 GTCAAAGAGTTATAGATTCCTCCTGATTTACCTGGGATCTCATTTGAATGTGAGAAGTTC 1304
Db 362 GGTCAAGAGTTATAGATTCCTCCTGATTTACCTGGGATCTCATTTGAATGTGAGAAGTTC 421
Qy 1305 AATTAATGAGGCTATTACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTGA 1364
Db 422 AATTAATGAGGCTATTACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTGA 481
Qy 1365 CAACTTTATCCGAGACACTTCTGACACTGGAGAGAGAACTGTGGTGTGAGTTATGGGG 1424
Db 482 CAACTTTATCCGAGACACTTCTGACACTGGAGAGAGAACTGTGGTGTGAGTTATGGGG 541
Qy 1425 TTTTCTCTTGAAATGCAGTCTTTGGACCACTGCTTAACCTAGNAGGCACTTAATCAGTG 1484
Db 542 TTTTCTCTTGAAATGCAGTCTTTGGACCACTGCTTAACCTAGNAGGCACTTAATCAGTG 601
Qy 1485 ACCAGCTAGATCTTGTCTGTGTCGCAAAAGTTATACATCCATCTTTTGAAGCAGTCTTCTCA 1544

Db 602 ACCACGTAGATCTTGCTGTCGCAAAAAGTTATACATCACTTTGAAGCAGTTTCTCA 661
Qy 1545 ATATGATCATGTAATTTGTCATAGACAGAAATTTGGAGTACTCTGATGGAAACCCCTGACTA 1604
Db 662 ATATGATCATGTAATTTGCCATAGACAGAAATTTGGAGTACTCTGATGGAAACCCCTGACTA 721
Qy 1605 AATGACAGAAACACAGCCTTAATGGATCCCTAGCTCAATATGCTCTTAATAGTATGGAA 1664
Db 722 AATGACAGAAACACAGCCTTAATGGATCCCTAGCTCAATATGCTCTTAATAGTATGGAA 781
Qy 1665 GTGTGAAGATGTGTGACATGCAAAAATTTGGCTCCCAACATTTTTTTCATCTTATTCCTTA 1724
Db 782 GTGTGAAGATGTGTGACATGCAAAAATTTGGCTCCCAACATTTTTTTCATCTTATTCCTTA 841
Qy 1725 ATGAAGCAACAATCTGCAGCTGCTCTCTGCAACAGAAATGACACTGCTGGGATGAGC 1784
Db 842 ATGAAGCAACAATCTGCAGCTGCTCTCTGCAACAGAAATGACACTGCTGGGATGAGC 901
Qy 1785 TGAAGCTCAGTGTGACAGGCACTGATGCTCTGAGTTAACTCAAGGAACCATGGG 1844
Db 902 TGAAGCTCAGTGTGACAGGCACTGATGCTCTGAGTTAACTCAAGGAACCATGGG 961
Qy 1845 CTTTGTGGAGGTGGCTGTACTGAAACTCATTTGGCTGATATATCAGACACAAGACTC 1904
Db 962 CTTTGTGGAGGTGGCTGTACTGAAACTCATTTGGCTGATATATCAGACACAAGACTC 1021
Qy 1905 ACAACGACCAGAAAGCATTTCTCAAGATGATGAATGTACTCAACAGAACTTCAATTA 1964
Db 1022 ACAACGACCAGAAAGCATTTCTCAAGATGATGAATGTACTCAACAGAACTTCAATTA 1081
Qy 1965 TTGCTGAAGCAATTTGCAAGTCCCTAGATCTGTTGCTGCTTTTGAACATGATGGAG 2024
Db 1082 TTGCTGAAGCAATTTGCAAGTCCCTAGATCTGTTGCTGCTTTTGAACATGATGGAG 1141
Qy 2025 GTCAAAATTCCTAGTACATGAAGTATGGACACCTTTGGTTCAGTTCAGGAGATTTCCCT 2084
Db 1142 GTCAAAATTCCTAGTACATGAAGTATGGACACCTTTGGTTCAGTTCAGGAGATTTCCCT 1201
Qy 2085 GTGTTGCTAACTGGCCAGATTTGCTTTTCAAGTGTGCTGTGATTAACAATAGCCAGG 2144
Db 1202 GTGTTGCTAACTGGCCAGATTTGCTTTTCAAGTGTGCTGTGATTAACAATAGCCAGG 1261
Qy 2145 AAGAACTCACTGGTCTTTTGAAGACACACGTGTCATTTTGGTCCCAAGCTGCC 2204
Db 1262 AAGAACTCACTGGTCTTTTGAAGACACACGTGTCATTTTGGTCCCAAGCTGCC 1321
Qy 2205 TTCCACATGAAGCTGTGGGCTCAGCCAGCAACCTGACCTTGGACTGTTGACTGCAAGC 2264
Db 1322 TTCCACATGAAGCTGTGGGCTCAGCCAGCAACCTGACCTTGGACTGTTGACTGCAAGC 1381
Qy 2265 TTAGTGGCTTACAGGTGGCTGTAGACAGACCCAAATTTGATTTGGGATCTTTTCATATGTTA 2324
Db 1382 TTAGTGGCTTACAGGTGGCTGTAGACAGACCCAAATTTGATTTGGGATCTTTTCATATGTTA 1441
Qy 2325 TTGAAGATAAACTAAGAAATGACATGTTCTGATTAACAGAAACAAATAAATAGT 2384
Db 1442 TTGAAGATAAACTAAGAAATGACATGTTCTGATTAACAGAAACAAATAAATAGT 1501
Qy 2385 CTGTGGCAATTCAGAAAAAATTT 2419
Db 1502 CTGTGGCAATTCAGAAAAAATTT 1536

RESULT 8
ADP43721
ID ADP43721 standard; cDNA; 1546 BP.
AC ADP43721;
XX
XX
DT 18-NOV-2004 (first entry)
XX
DE Human PMMM-37 encoding cDNA SEQ ID NO:78.

XX
KW human; protein modification and maintenance molecule; PMMM;
KW gastrointestinal; cardiovascular; immunosuppressive; anti-inflammatory;
KW cytostatic; neuroprotective; gynaecological; gene therapy;
KW gastrointestinal disorder; cardiovascular disorder; autoimmune disorder;
KW inflammatory disorder; cell proliferative disorder;
KW developmental disorder; epithelial disorder; neurological disorder;
KW reproductive disorder; gene; ss; single nucleotide polymorphism; SNP.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 14..1510
FT /*tag= a
FT /product= "PMMM-37"
FT /replace(541,C)
FT /*tag= b
FT /*tag= c
FT variation
FT /*standard_name= "single nucleotide polymorphism (SNP)"
FT replace(547,T)
FT /*tag= c
FT /*standard_name= "single nucleotide polymorphism (SNP)"
XX
PN WO2004053068-A2.
XX
PD 24-JUN-2004.
XX
PF 03-DEC-2003; 2003WO-US038573.
XX
PR 05-DEC-2002; 2002US-0431639P.
PR 17-DEC-2002; 2002US-0434315P.
PR 24-JAN-2003; 2003US-0442442P.
PR 30-JAN-2003; 2003US-0444141P.
PR 21-FEB-2003; 2003US-0449491P.
XX
PA (INCY-) INCYTE CORP.
XX
PI Becha SD, Hafalia AJA, Swarnakar A, Ramkumar J, Richardson TW;
PI Kabie AB, Marquis JP, Khare R, Lee SY, Tran UK, Bhatia UG;
PI Burrill JD, Blake JU, Ho A, Zheng W, Gao J, Chawla NK, Mason PM;
PI Jin P, Lee S;
XX
DR WPI; 2004-468843/44.
DR P-PSDB; ADP43680.
XX
PT New human protein modification and maintenance molecules and
PT polynucleotides for diagnosing, preventing or treating diseases
PT associated with aberrant protein expression, e.g. cardiovascular or cell
PT proliferative disorders.
XX
PS Claim 5; SEQ ID NO 78; 268pp; English.
XX
CC The present sequence encodes a human protein modification and maintenance
CC molecule (PMMM) (1), designated PMMM-37. Also described: (1) an isolated
CC polynucleotide encoding (1); (2) a recombinant polynucleotide comprising
CC a promoter sequence operably linked to the polynucleotide in (1); (3) a
CC cell transformed with the recombinant polynucleotide; (4) a transgenic
CC organism comprising the recombinant polynucleotide; (5) methods of
CC producing or purifying (1); (6) an isolated antibody that specifically
CC binds to (1); (7) detecting a target polynucleotide or (1) in a sample;
CC (8) compositions comprising the polypeptide, an agonist compound, an
CC antagonist compound or an antibody, and an excipient; (9) treating
CC diseases or conditions associated with decreased expression or
CC overexpression of functional human PMMM; (10) screening for a compound
CC that is effective as an agonist or antagonist of (1), that specifically
CC binds to (1), that modulates the activity of (1), or is effective in
CC altering expression of the target polynucleotide; (11) screening for
CC potential toxicity of a test compound; (12) a diagnostic test for a
CC condition or disease associated with the expression of PMMM in a
CC biological sample; (13) diagnosing a condition or disease associated with
CC the expression of PMMM in a subject; (14) preparing a polyclonal or
CC monoclonal antibody with the specificity of the antibody in (6); (15) a
CC polyclonal or monoclonal antibody produced by the method in (14); (16)
CC compositions comprising the polyclonal or monoclonal antibody, and a

CC carrier; (17) generating an expression profile of a sample containing the
 CC polynucleotides; and (18) an array comprising different nucleotide
 CC molecules affixed at distinct physical locations on a solid substrate,
 CC where at least one nucleotide molecule comprises a first oligonucleotide
 CC or polynucleotide sequence specifically hybridizable with at least 30
 CC contiguous nucleotides of the target polynucleotide. PMM sequences have
 CC gastrointestinal, cardiovascular, immunosuppressive, antiinflammatory,
 CC cytostatic, neuroprotective and gynaecological activities, and can be
 CC used in gene therapy. The composition and methods are useful for
 CC diagnosing, preventing or treating diseases or conditions associated with
 CC aberrant expression of PMM, such as gastrointestinal, cardiovascular,
 CC autoimmune/inflammatory, cell proliferative, developmental, epithelial,
 CC neurological or reproductive disorders. They may also be used for
 CC assessing the effects of exogenous compounds on the expression of nucleic
 CC acid and amino acid sequences of PMM. The PMM or its fragments are also
 CC useful in screening compounds for effectiveness as agonist or antagonist
 CC of the polypeptides, or in altering the expression of the target
 CC polynucleotide and compounds that specifically bind to or modulate the
 CC activity of the polypeptide. The microarray is useful in monitoring or
 CC measuring protein-protein interactions, drug-target interactions, and
 CC gene expression profiles.

XX Sequence 1546 BP; 445 A; 327 C; 331 G; 443 T; 0 U; 0 Other;

Query Match 53.2%; Score 1314.2; DB 13; Length 1546;
 Best Local Similarity 87.6%; Pred. No. 1.6e-207;
 Matches 1542; Conservative 0; Mismatches 3; Indels 216; Gaps 1;

QY 617 ACCTGAAGTAAGATGTCGTGTTGGAAGCTAAGAGCCATCAATGTGTAAAGTGAACCA 676
 DB 2 ACCTGAAGTAAGATGTCGTGTTGGAAGCTAAGAGCCATCAATGTGTAAAGTGAACCA 61
 QY 677 CTGACAACTGAGAGCTCAGGACCACTTCTGCTCTGAAAGAAATTCATCATCATGC 736
 DB 62 CTGACAACTGAGAGCTCAGGACCACTTCTGCTCTGAAAGAAATTCATCATCATGC 121
 QY 737 TATGGCCCTCAGGTAGGCTGAAGCAGCTGCACAAATGGCTTTGGAGTTACGTGTACA 796
 DB 122 TATGGCCCTCAGGTAGGCTGAAGCAGCTGCACAAATGGCTTTGGAGTTACGTGTACA 181
 QY 797 ACCTCAGTCTCAGCTCTGCTCAGTCACTTTTGGTCAACATCCCATTTTAAAGATC 856
 DB 182 ACCTCAGTCTCAGCTCTGCTCAGTCACTTTTGGTCAACATCCCATTTTAAAGATC 241
 QY 857 CTGACAGCTCCATACAGATCATGTCTCAAGCTTCAGTGTGCTTATTCACAGCT 916
 DB 242 CTGACAGCTCCATACAGAAATCATGTCTCAAGCTTCAGTGTGCTTATTCACAGCT 301
 QY 917 ATTCTTTGCTGCAACCTGATTTGAAAATGTTACAGATTAGGCTTTGACCCACCACTGTC 976
 DB 302 ATTCTTTGCTGCAACCTGATTTGAAAATGTTACAGATTAGGCTTTGACCCACCACTGTC 361
 QY 977 ATTAGATTAAATAAATCTTTTGAATCTTTGATCACTTTTCAAGTCTGATACCTGT 1036
 DB 362 ATTAGATTAAATAAATCTTTTGAATCTTTGATCACTTTTCAAGTCTGATACCTGT 421
 QY 1037 GGTGTGCAATCCAGTGACCTTTAGTACTCTCAGATCCTCTCTTTGTTGGTGGTAGT 1096
 DB 422 GGTGTGCAATCCAGTGACCTTTAGTACTCTCAGATCCTCTCTTTGTTGGTGGTAGT 481
 QY 1097 ATATTAAACAAGTAAACCTGCTGTATGCTCACAGAAAGAAACAGAGCATGTCAGTGTCT 1156
 DB 482 ATATTAAACAAGTAAACCTGCTGTATGCTCACAGAAAGAAACAGAGCATGTCAGTGTCT 541
 QY 1157 TTGATCCTGAGAGCTTTTGTCTTACAAATTCAGAAAAATGCTGAAGGCCACATCAATTTTA 1216
 DB 542 TTGATCCTGAGAGCTTTTGTCTTACAAATTCAGAAAAATGCTGAAGGCCACATCAATTTTA 601
 QY 1217 GGAAGAGTTAATCTGACTTTTAAAGGTCAAGAGTTATAGATTCACCTGTATACCT 1276
 DB 602 GGAAGAGTTAATCTGACTTTTAAAGGTCAAGAGTTATAGATTCACCTGTATACCT 661
 QY 1277 GGGATACCTCATTGAAATGTGAGAAGTTCAATTAATGAGGCTATTACCTATCAAAAAATCA 1336

DB 662 GGGATACCTCATTGAAATGTGAGAAGTTCAATTAATGAGGCTATTACCTATCAAAAAATCA 721
 QY 1337 ACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGACACTTCTGACACTGGA 1396
 DB 722 ACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGACACTTCTGACACTGGA 781
 QY 1397 GAAGGAACCTGTGCTGCTCAGTTATGGGGTTTCTCTTGAATAATGCACTTGTGACCAAGCTG 1456
 DB 782 GAAGGAACCTGTGCTGCTCAGTTATGGGGTTTCTCTTGAATAATGCACTTGTGACCAAGCTG 841
 QY 1457 CTTAAACCTAGGAAGGAGCACTAATCAGTGAACCACTAGTCTCTGTGTCGCAAAAGTT 1516
 DB 842 CTTAAACCTAGGAAGGAGCACTAATCAGTGAACCACTAGTCTCTGTGTCGCAAAAGTT 901
 QY 1517 ATACATCCATCTTTGAAGCAGTTTCTCAATATGATCGTATTTATTGCCCATAGACAAAT 1576
 DB 902 ATACATCCATCTTTGAAGCAGTTTCTCAATATGATCGTATTTATTGCCCATAGACAAAT 961
 QY 1577 GGAGTGACTCTGATGGAACCCCTGACTAAATGACAGGAACACAGCCTATTGGATCCCTA 1636
 DB 962 GGAGTGACTCTGATGGAACCCCTGACTAAATGACAGGAACACAGCCTATTGGATCCCTA 1021
 QY 1637 GGCCTCAATATGTCCTAATAGTTATGGAAGTGTGAAAGATGTGTGCACTGCAAAATTTGCG 1696
 DB 1022 GGCCTCAATATGTCCTAATAGTTATGGAAGTGTGAAAGATGTGTGCACTGCAAAATTTGCG 1081
 QY 1697 TCCCAACATTTTTTTCATCTTATCTTAATGAAGCAACATCTGACGCTGCTTCTCTGC 1756
 DB 1082 TCCCAACATTTTTTTCATCTTATCTTAATGAAGCAACATCTGACGCTGCTTCTCTGC 1141
 QY 1757 AACAGAAATGACACTGCTGGGATGAGCTGAAGCTCAGTGTGACAGCGCACTGCAATGTC 1816
 DB 1142 AACAGAAATGACACTGCTGGGATGAGCTGAAGCTCAGTGTGACAGCGCACTGCAATGTC 1201
 QY 1817 CTGCAAGTTAAACCTCAAGGAACCATGGGCTTTGTTGGAGGTGGCTGTACTGAAACTCAT 1876
 DB 1202 CTGCAAGTTAAACCTCAAGGAACCATGGGCTTTGTTGGAGGTGGCTGTACTGAAACTCAT 1261
 QY 1877 TTGGCTGATATATCAGACACAAAGACTCACAACGACCCCAAGAGCATTTCTAAAGATGAT 1936
 DB 1262 TTGGCTGATATATCAGACACA ----- 1283
 QY 1937 GAATGCTACTCAACAGAACTTCAATTAATGCTGAAGCATTTTTCAGTGGCCCTAGAACTCT 1996
 DB 1284 ----- 1283
 QY 1997 GTTGTGGCTCTTTAGAACATGATGGAGGTGAAATTTCTCACTGACATGAAGTATGGACAC 2056
 DB 1284 ----- 1283
 QY 2057 CTTTGGTCAGTTGAGGAGATTTCTCCCTGTGTGTCTAACTGGCCAGATTTGCTTTTACAG 2116
 DB 1284 -----AG 1285
 QY 2117 TGTGCTGTGGATTATACAATAGCAGGAAGAACTCAACTGCTGCTTTCTTAAGAAAGCACA 2176
 DB 1286 TGTGCTGTGGATTATACAATAGCAGGAAGAACTCAACTGCTGCTTTCTTAAGAAAGCACA 1345
 QY 2177 CGTCTGCTCAATTTGTGCCCAAAAGCTGCTTCCACATGAAGCTGTGGGCTCAGCCAGCAAC 2236
 DB 1346 CGTCTGCTCAATTTGTGCCCAAAAGCTGCTTCCACATGAAGCTGTGGGCTCAGCCAGCAAC 1405
 QY 2237 CTGACCTTGGACTGTTTGAAGCTTAGTGCCCTACAGGTGGCTGTAGAGACAGCC 2296
 DB 1406 CTGACCTTGGACTGTTTGAAGCTTAGTGCCCTACAGGTGGCTGTAGAGACAGCC 1465
 QY 2297 AATTGATTTGGGATCTTTTCATCTTATGAAGATAAACTAAGAGAAATAGCATGTTC 2356
 DB 1466 AATTGATTTGGGATCTTTTCATCTTATGAAGATAAACTAAGAGAAATAGCATGTTC 1525
 QY 2357 GTATTACAGAGAAAAAATA 2377

Db 1526 GTATTACAGAGAAACAATA 1546

RESULT 9

AAC99977

ID AAC99977 standard; cDNA; 1274 BP.

XX AAC99977;

XX 13-MAR-2001 (first entry)

XX Human secreted protein gene 142 SEQ ID NO:170.

XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
XX antirheumatic; antiproliferative; cytostatic; antibiotic; vasotropic;
XX cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
XX fungicide; ophthalmologic; gene therapy; pathological condition;
XX autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
XX neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
XX cerebrovascular disorder; angiogenesis; nervous system disorder;
XX Alzheimer's disease; infection; ocular disorder; corneal infection;
XX wound healing; skin aging; food additive; preservative; chromosome 20;
XX ss.

XX Homo sapiens.

XX WO200070042-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-US012788.

XX 13-MAY-1999; 99US-0134068P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;
XX Duan RD, Florence KA, Soppet DR;

XX WPI; 2000-679828/66.

XX P-PSDB; AAB56236.

XX Isolated nucleic acid molecule encoding a human secreted protein is used
XX in preventing, treating or ameliorating a medical condition.

XX Claim 1; Page 939-940; 1065pp; English.

XX The polynucleotide sequences given in AAC99818 to AAC99977 encode the
XX human secreted proteins given in AAB56077 to AAB56362. Human secreted
XX proteins have activities based on the tissues and cells the genes are
XX expressed in. Examples of activities include: immunosuppressive;
XX antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiac;
XX vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
XX virucide; fungicide; and ophthalmologic. The human secreted
XX polynucleotides and proteins can be used to prevent, treat or ameliorate
XX a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
XX dogs, chickens or sheep. They are also used in diagnosing a pathological
XX condition or susceptibility to a pathological condition. Disorders which
XX are diagnosed or treated include autoimmune diseases e.g. rheumatoid
XX arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
XX liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
XX disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
XX e.g. Alzheimer's disease, infections caused by bacteria, viruses and
XX fungi and ocular disorders e.g. corneal infection. The proteins can also
XX be used to aid wound healing and epithelial cell proliferation, to
XX prevent skin aging due to sunburn, to maintain organs before
XX transplantation, for supporting cell culture of primary tissues, to
XX regenerate tissues and in chemotaxis. The proteins can also be used as a
XX food additive or preservative to increase or decrease storage
XX capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
XX in the exemplification of the present invention

XX Sequence 1274 BP; 390 A; 256 C; 271 G; 357 T; 0 U; 0 Other;

Query Match		50.7%;	Score 1253.8;	DB 3;	Length 1274;
Best Local Similarity		99.8%;	Pred. No. 1.5e-197;		
Matches 1266;		Conservative	0;	Mismatches	2;
				Indels	1;
				Gaps	1;
QY	1156	TTTGTATCCTGAGAGCCTTTTTCCTTACAAATTCAGAAAAATGCTGAAGGCGCACATCATTTT	1215		
DB	1	TTTGTATCCTGAGAGCCTTTTTCCTTACAAATTCAGAAAAATGCTGAAGGCGCACATCATTTT	60		
QY	1216	AGGAAAGAGTAAATTTGTACCTTTTAAAGGTCAAGAGGTTATAGATTTCACCTGTATTACC	1275		
DB	61	AGGAAAGAGTAAATTTGTACCTTTTAAAGGTCAAGAGGTTATAGATTTCACCTGTATTACC	120		
QY	1276	TGGGATACCTATGAAAATGTCAAGAGTTCAATTAATGAGGCTATTACCTATCAAAAAATC	1335		
DB	121	TGGGATACCTATGAAAATGTCAAGAGTTCAATTAATGAGGCTATTACCTATCAAAAAATC	180		
QY	1336	AACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGACACTCTTGACACTGG	1395		
DB	181	AACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGACACTCTTGACACTGG	240		
QY	1396	AGAAAGAACTGTGGTGGTCAGTTATGGGGTTTCTCTTGAAAAATGCAGTCTTGACCACT	1455		
DB	241	AGAAAGAACTGTGGTGGTCAGTTATGGGGTTTCTCTTGAAAAATGCAGTCTTGACCACT	300		
QY	1456	GCTTAACCTAGGAAGGCAGCTAATCAGTGACCACTAGATCTTGCTGTGCCAAAAAGT	1515		
DB	301	GCTTAACCTAGGAAGGCAGCTAATCAGTGACCACTAGATCTTGCTGTGCCAAAAAGT	360		
QY	1516	TATACATCCATCTTTGAAGCAGTTTCTCAATATGCATCGTATATTATGCCATAGACAGAAT	1575		
DB	361	TATACATCCATCTTTGAAGCAGTTTCTCAATATGCATCGTATATTATGCCATAGACAGAAT	420		
QY	1576	TGGAGTGACTCTGATGGAAACCCCTGACTAAATAGACAGAAACACAGCCCTATTGGATCCCT	1635		
DB	421	TGGAGTGACTCTGATGGAAACCCCTGACTAAATAGACAGAAACACAGCCCTATTGGATCCCT	480		
QY	1636	AGGCTCAATATGCTCTTAATAGTATGGAAGTGTGAAGATGTGACACTGCAAAATTTGG	1695		
DB	481	AGGCTCAATATGCTCTTAATAGTATGGAAGTGTGAAGATGTGACACTGCAAAATTTGG	540		
QY	1696	CTCCCAACATTTTTTTCATCTTATTCCTAAT- GAAGCAACAATCTGCAGCTTGCTCTCT	1754		
DB	541	CTCCCAACATTTTTTTCATCTTATTCCTAATGGAAGCAACAATCTGCAGCTTGCTCTCT	600		
QY	1755	GCAACAGAAATGACACTGCTGGGATGAGCTGAAAGTCAAGTGTGACAGCGCACTGCATG	1814		
DB	601	GCAACAGAAATGACACTGCTGGGATGAGCTGAAAGTCAAGTGTGACAGCGCACTGCATG	660		
QY	1815	TCTTGCAGTTAATCACTCAGGAACCAATGGGCTTTGTTGGAGGTGGCTGTACTGAACTC	1874		
DB	661	TCTTGCAGTTAATCACTCAGGAACCAATGGGCTTTGTTGGAGGTGGCTGTACTGAACTC	720		
QY	1875	ATTGGCTCATATATCAGACACAGACTCAACAGACCAGAAAGCATTTCTCAAAAGATG	1934		
DB	721	ATTGGCTCATATATCAGACACAGACTCAACAGACCAGAAAGCATTTCTCAAAAGATG	780		
QY	1935	ATGAATGTACTCAACAGAACTTCAATTAATTTGCTGAAGCATTTTGCAGTGCCTTGAAT	1994		
DB	781	ATGAATGTACTCAACAGAACTTCAATTAATTTGCTGAAGCATTTTGCAGTGCCTTGAAT	840		
QY	1995	CTGTTGTTGGCTTTTGTAGAACATGATGGAGGTGAATTTCTCACTGCATGATGATGGAC	2054		
DB	841	CTGTTGTTGGCTTTTGTAGAACATGATGGAGGTGAATTTCTCACTGCATGATGATGGAC	900		
QY	2055	ACCTTTGGTTCAGTTTCAGGAGATTTCTCCCTGTGTTGCTAACTGGCCAGATTTGCTTTTAC	2114		
DB	901	ACCTTTGGTTCAGTTTCAGGAGATTTCTCCCTGTGTTGCTAACTGGCCAGATTTGCTTTTAC	960		
QY	2115	AGTGTGGCTGTGGATTTATCAATAGCCAGGAAGAACTCAACTGGTCTTTCTTTAAGAGCA	2174		
DB	961	AGTGTGGCTGTGGATTTATCAATAGCCAGGAAGAACTCAACTGGTCTTTCTTTAAGAGCA	1020		

QY 2175 CAGCTGCTCCATTGTCGCCAAGCTGCTTCCATGAGCTGTGGGCTCAGCCAGCA 2234
DB 1021 CAGCTGCTCCATTGTCGCCAAGCTGCTTCCATGAGCTGTGGGCTCAGCCAGCA 1080
QY 2235 ACCTGACCTTTGGACTGTTTGCACGCAAGCTTGTGCGCTTACAGGCTGTAGACACAG 2294
DB 1081 ACCTGACCTTTGGACTGTTTGCACGCAAGCTTGTGCGCTTACAGGCTGTAGACACAG 1140
QY 2295 CCAATTGATTTGGGATCTTTTCATATGTTATTGAAGATAAACTAAGAGATAGCATGT 2354
DB 1141 CCAATTGATTTGGGATCTTTTCATATGTTATTGAAGATAAACTAAGAGATAGCATGT 1200
QY 2355 TCGTATTACAGAGAAACAATAAATAGCTGCTGTGGCAATTGAGAAAAA 2414
DB 1201 TCGTATTACAGAGAAACAATAAATAGCTGCTGTGGCAATTGAGAAAAA 1260
QY 2415 AAAAAAAA 2423
DB 1261 AAAAAAAA 1269

RESULT 10
ABK35896
ID ABK35896 standard; cDNA; 694 BP.
XX
AC ABK35896;
XX
DT 08-MAY-2002 (first entry)
XX
DE cDNA sequence #287 encoding novel human secreted protein.
XX
KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;
KW immune deficiency disorder; blood disorder; inflammatory disorder;
KW infectious disorder; allergic condition; neurodegenerative disorder;
KW liver fibrosis; coagulation disorder; gene therapy; antimicrobial;
KW tumour; cancer; hepatotropic; immunosuppressive; antirheumatic; gene; ss.
OS Homo sapiens.
XX
XX WO200177289-A2.
XX
XX 18-OCT-2001.
XX
XX 29-MAR-2001; 2001WO-US010232.
XX
XX 06-APR-2000; 2000US-0195605P.
XX
XX (GEMY) GENETICS INST INC.
XX
XX Jacobs K, Mccoy JM, Lavallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M, Agostino MJ, Bowman MR, Spaulding V, Wong GG;
XX Clark HF, Fechtel K, Howes SH, Resnick RJ, Gulukota K, Graham JR;
XX WPI; 2002-179322/23.
XX

PT Six hundred and twenty three polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for treating
PT immune deficiencies and disorders such as autoimmune disorders.
XX
XX Claim 1; Page 248; 393pp; English.
XX
XX The present invention relates to the isolation of novel cDNA sequences
XX which encode human secreted proteins. The cDNA sequences have been
XX derived from a variety of human tissues. The invention also provides a
XX method for producing proteins from these polynucleotide sequences. The
XX proteins are useful for identifying compounds that modulate their
XX activity and production. The sequences of the invention are useful for
XX treating diseases such as hyperproliferative disorders (e.g. cancer),
XX immune deficiency disorders (e.g. severe combined immunodeficiency
XX (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders
XX (e.g. thrombocytopaenia), inflammatory disorders (e.g. arthritis),
XX infectious disorders (e.g. hepatitis), allergic conditions (e.g. asthma),
XX neurodegenerative disorders (e.g. Alzheimer's disease), liver fibrosis,
XX

CC coagulation disorders (e.g. haemophilia), and tumours. The polynucleotide
CC sequences of the invention are also useful in gene therapy. ABK35610-
CC ABK36232 represent the cDNA sequences of the invention that encode for
CC novel human secreted proteins
XX
SQ Sequence 694 BP; 215 A; 118 C; 160 G; 201 T; 0 U; 0 Other;
Query Match 24.0%; Score 593; DB 6; Length 694;
Best Local Similarity 100.0%; Pred. No. 1.2e-88;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TATGAGCCTTCGGAACCTTGTGGAGAGACTACAAAGTTTGGTTGTTATGCTCCCTTAGT 60
DB 102 TATGAGCCTTCGGAACCTTGTGGAGAGACTACAAAGTTTGGTTGTTATGCTCCCTTAGT 161
QY 61 TGGGCTCATACATTTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 120
DB 162 TGGGCTCATACATTTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 221
QY 121 AAACGACGACATTTCCCTGAGCAAGATAGTCTGGGACTTTTCAAAATCTTCAGAAAGCCAAAT 180
DB 222 AAACGACGACATTTCCCTGAGCAAGATAGTCTGGGACTTTTCAAAATCTTCAGAAAGCCAAAT 281
QY 181 CCAGGGGAAAGTAGCAGGCTTGGCAATCTTCAGGTAAGAAAGCAGCTTTGAAATCTGAGCTTC 240
DB 282 CCAGGGGAAAGTAGCAGGCTTGGCAATCTTCAGGTAAGAAAGCAGCTTTGAAATCTGAGCTTC 341
QY 241 ATATCGAAGAGAGATGAAATACCAAGTTCGATAGAAAGACCTGGCTTCTTGTAGCT 300
DB 342 ATATCGAAGAGAGATGAAATACCAAGTTCGATAGAAAGACCTGGCTTCTTGTAGCT 401
QY 301 GGGATATCTTTTCATAGGTCTCCATCTTGGAACATATCTTTTGCAGAGGTCTGCAAGCAG 360
DB 402 GGGATATCTTTTCATAGGTCTCCATCTTGGAACATATCTTTTGCAGAGGTCTGCAAGCAG 461
QY 361 TCTGTAATAATTTTCAGTCTCAAGGCAAAACAAAGAGATTTGAAAGAGTAAAGTAAATPAAAT 420
DB 462 TCTGTAATAATTTTCAGTCTCAAGGCAAAACAAAGAGATTTGAAAGAGTAAAGTAAATPAAAT 521
QY 421 ATTTGGAATTTACTAATTTTGTCAATTAATCAATCTTATGCTGATAGCTTCATAAATTCGA 480
DB 522 ATTTGGAATTTACTAATTTTGTCAATTAATCAATCTTATGCTGATAGCTTCATAAATTCGA 581
QY 481 ACTTTTGTATTTATAGCCACAATGCTGCATATTCATATCTTTAAATTCCTTAAAGATAAT 540
DB 582 ACTTTTGTATTTATAGCCACAATGCTGCATATTCATATCTTTAAATTCCTTAAAGATAAT 641
QY 541 TTTAATGTTAAACCTGATTAATGCAATAAATAGAAAAATGTGTTTACAAAAT 593
DB 642 TTTAATGTTAAACCTGATTAATGCAATAAATAGAAAAATGTGTTTACAAAAT 694
RESULT 11
AAK97580
ID AAK97580 standard; DNA; 458 BP.
XX
AC AAK97580;
XX
DT 13-SEP-1999 (first entry)
XX
DE Extended human secreted protein coding sequence, SEQ ID NO. 44.
XX
KW Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW cellular differentiation; immune system regulator; anti-inflammatory;
KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW genetic disease; ss.
OS Homo sapiens.
XX
XX WO9931236-A2.
XX
XX 24-JUN-1999.

SEQUENCE COMPARISON (A)

XX PF 17-DEC-1998; 98WO-IB0002122.
XX PR 17-DEC-1997; 97US-0069957P.
XX PR 09-FEB-1998; 98US-0074121P.
XX PR 13-APR-1998; 98US-0081563P.
XX PR 10-AUG-1998; 98US-0096116P.
XX (GEST) GENSET.
XX PA Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX PI WPI; 1999-385906/32.
XX DR P-PSDB; AAY35896.
XX DR New isolated human secreted proteins.
XX PT Claim 1; Page 179; 516pp; English.
XX PS This sequence represents an extended human secreted protein coding
XX CC sequence of the invention. The secreted proteins can be used in treating
XX CC or controlling a variety of human conditions. The secreted proteins may
XX CC act as cytokines or may affect cellular proliferation or differentiation
XX CC or may act as immune system regulators, haematopoiesis regulators, tissue
XX CC growth regulators, regulators of reproductive hormones or cell movement
XX CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
XX CC tumour inhibition activity. The DNAs can be used in forensic procedures
XX CC to identify individuals or in diagnostic procedures to identify
XX CC individuals having genetic diseases resulting from abnormal expression of
XX CC the genes corresponding to the extended cDNAs. They are also useful for
XX CC constructing a high resolution map of the human chromosomes. They can
XX CC also be used for gene therapy to control or treat genetic diseases
XX
SQ Sequence 458 BP; 160 A; 74 C; 102 G; 121 T; 0 U; 1 Other;

Query Match 17.1%; Score 422.4; DB 2; Length 458;
Best Local Similarity 99.5%; Pred. No. 1.5e-60;
Matches 434; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 TATGAGCCTTCGGAACCTGTGGAGAGACTACAAAGTTTGGTGTATGGTCCCTTTAGT 60
DB 11 TATGAGCCTTCGGAACCTGTGGAGAGACTACAAAGTTTGGTGTATGGTCCCTTTAGT 70
QY 61 TGGGCTCATACATTTGGGCTGTGACAGATCAAAAGCAGCCCTGTTTCCAAATACCTAA 120
DB 71 TGGGCTCATACATTTGGGCTGTGACAGATCAAAAGCAGCCCTGTTTCCAAATACCTAA 130
QY 121 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTCAAATCTTCAGAAGAGCCAAAT 180
DB 131 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTCAAATCTTCAGAAGAGCCAAAT 190
QY 181 CCAGGGGAAG-TAGCAGGCTTGCATCTTCAGGTAAGAGAGCAGCTTGAATCTGAGCTT 239
DB 191 CCAGGGGAAGTAGCAGGCTTGCATCTTCAGGTAAGAGAGCAGCTTGAATCTGAGCTT 250
QY 240 CATATCGAAGAGAGATGAAAAATACCAAGTTGGATTAGAAAGAACTGGCTTCTTTAGC 299
DB 251 CATATCGAAGAGAGATGAAAAATACCAAGTTGGATTAGAAAGAACTGGCTTCTTTAGC 310
QY 300 TGGGATATCTTTCATAGGTCTCCATCTTGGACATACATTTTTCAGAGGTCTGCAAGCA 359
DB 311 TGGGATATCTTTCATAGGTGACCATCTTGGACATACATTTTTCAGAGGTCTGCAAGCA 370
QY 360 GTCTGTAATAATTTTCAGTCTCAAGCAAAACAAAAGAGATATTGAAGAGTGAAGTAAATAA 419
DB 371 GTCTGTAATAATTTTCAGTCTCAAGCAAAACAAAAGAGATATTGAAGAGTGAAGTAAATAA 430
QY 420 TATTGGAATTAATAA 435
DB 431 TATTGGAATTAATAA 446

ADP18847
ID ADP18847 standard; cDNA; 458 BP.
XX AC ADP18847;
XX DT 26-AUG-2004 (first entry)
XX DE Human secreted polynucleotide #103.
XX KW Human; secreted protein; gene; ss; genetic disease.
XX OS Homo sapiens.
XX PN US2004110939-A1.
XX PD 10-JUN-2004.
XX PF 15-OCT-2001; 2001US-00978360.
XX PR 17-DEC-1998; 98WO-IB002122.
XX PR 09-FEB-1999; 99WO-IB000282.
XX PR 21-JUN-2000; 2000WO-IB000951.
XX PR 15-SEP-2000; 2000US-00663600.
XX (GEST) GENSET SA.
XX Dumas Milne Edwards J, Bougueleret L, Jobert S, Clusel C;
XX Duclert A;
XX WPI; 2004-440404/41.
XX DR P-PSDB; ADP19252.
XX New isolated polynucleotide encoding secreted polypeptide, useful for
XX gene therapy, or in diagnostic procedures to identify individuals having
XX genetic diseases resulting from abnormal expression of the genes.
XX Claim 1; SEQ ID NO 103; 113pp; English.
XX The invention relates to human cDNA sequences that encode human secreted
XX proteins. The invention also relates to an antibody that specifically
XX binds to a polypeptide of the invention and a method of binding the
XX polypeptide to an antibody. The polynucleotides are useful for expressing
XX the entire secreted proteins which they encode and for distinguishing
XX human tissues and cells from non-human tissues and cells, and for
XX distinguishing between human tissues and cells that do or do not express
XX the polynucleotides comprising the cDNAs. The polynucleotides and
XX polypeptides are useful in forensic procedures or diagnostic procedures
XX to identify individuals with genetic diseases or resulting from abnormal
XX expression of the genes corresponding to the cDNAs. The sequences are
XX also useful in gene therapy to control or treat genetic diseases. This
XX sequence represents a human secreted polynucleotide of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification but was obtained in electronic format from USPTO at
XX seqdata.uspto.gov/sequence.html.
SQ Sequence 458 BP; 160 A; 74 C; 102 G; 121 T; 0 U; 1 Other;

Query Match 17.1%; Score 422.4; DB 12; Length 458;
Best Local Similarity 99.5%; Pred. No. 1.5e-60;
Matches 434; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 TATGAGCCTTCGGAACCTGTGGAGAGACTACAAAGTTTGGTGTATGGTCCCTTTAGT 60
DB 11 TATGAGCCTTCGGAACCTGTGGAGAGACTACAAAGTTTGGTGTATGGTCCCTTTAGT 70
QY 61 TGGGCTCATACATTTGGGCTGTGACAGATCAAAAGCAGCCCTGTTTCCAAATACCTAA 120
DB 71 TGGGCTCATACATTTGGGCTGTGACAGATCAAAAGCAGCCCTGTTTCCAAATACCTAA 130
QY 121 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTCAAATCTTCAGAAGAGCCAAAT 180
DB 131 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTCAAATCTTCAGAAGAGCCAAAT 190

QY 181 CCAGGGGAAG-TACAGGCTTGCATCTTTCAGGTAAAGAGCAGCTTTGAACTCTGAGCTT 239
DB 191 CCAGGGGAAGTAGCAGCTTGCATCTTTCAGGTAAAGAGCAGCTTTGAACTCTGAGCTT 250
QY 240 CATATCGAAGAGAGATGAATAACACCTTGGATTAGAAAGAACTGGCTTCTTGTAGC 299
DB 251 CATATCGAAGAGAGATGAATAACACCTTGGATTAGAAAGAACTGGCTTCTTGTAGC 310
QY 300 TGGGATATCTTTTCATAGGTGTCCATCTTGGAAACATACCTTTTTCAGAGGTCTGCAAGCA 359
DB 311 TGGGATATCTTTTCATAGGTGTCCATCTTGGAAACATACCTTTTTCAGAGGTCTGCAAGCA 370
QY 360 GTCTGTAATAATTCAGTCTCAAGCAAAACAAAGAGATGTAAGAGTGAAGTAAATAAA 419
DB 371 GTCTGTAATAATTCAGTCTCAAGCAAAACAAAGAGATGTAAGAGTGAAGTAAATAAA 430
QY 420 TATTTGGAATTACTAA 435
DB 431 TATTTGGAATTACTAA 446

RESULT 13
AAA77867/c
ID AAA77867 standard; cDNA; 421 BP.
XX
AC AAA77867;
DT 14-NOV-2000 (first entry)
XX
DE cDNA encoding human colon tumour polypeptide, SEQ ID NO:147.
XX
KW Human colon tumour polypeptide; tumour antigen; cancer; vaccine;
KW immunotherapy; diagnosis; progression; ss.
OS Homo sapiens.
XX
PN WO200037643-A2.
XX
XX 29-JUN-2000.
XX
XX 23-DEC-1999; 99WO-US030909.
XX
PR 23-DEC-1998; 98US-00221298.
PR 02-JUL-1999; 99US-00347496.
PR 22-SEP-1999; 99US-00401064.
PR 19-NOV-1999; 99US-00444242.
PR 02-DEC-1999; 99US-00454150.
XX
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk J;
PI Wang T, Yuqiu J;
PI
XX WPI; 2000-442671/38.
XX
XX New colon tumor polypeptides used to inhibit the development of cancer,
PT especially colon cancer, and for diagnosing and monitoring the
PT progression of the cancer.
XX
XX
PS Claim 29; Page 133; 229pp; English.
XX

Sequences AAA7722-A78199 represent 478 cDNAs encoding proteins or
CC portions of proteins which are associated with human colon tumours. The
CC invention also specifically discloses 8 human colon tumour proteins.
CC (AAB11897-B11904). The nucleic acids, the polypeptides they encode, and
CC antigen presenting cells (APCs, preferably dendritic cells) expressing
CC such polypeptides may be used in vaccines that target tumour cells,
CC especially colon tumour cells, thereby inhibiting the development of
CC cancer. T-cells specific for the polypeptide expressed by the APC are
CC used to remove tumour cells from biological samples, especially blood or
CC fractions thereof. The sample or the isolated T-cells specific for the
CC polypeptide can then be used to inhibit cancer development. CD4+ and/or
CC CD8+ T-cells from a patient may be incubated with a polypeptide or

CC nucleic acid of the invention, or an APC expressing such a polypeptide,
CC to cause the proliferation of specific T-cells. The T-cells can be cloned
CC and then administered back to the patient to inhibit cancer development.
CC Nucleic acids encoding the polypeptides and antibodies against the
CC polypeptides may be used to determine the expression level of a tumour
CC protein of the invention, and therefore to determine whether cancer cells
CC are present. Such diagnostic methods may also be used to monitor the
CC progression of a cancer by repeating the processes at time intervals, and
CC comparing the current result to previous results. The present sequence
CC represents a cDNA encoding a human colon tumour polypeptide
XX
SQ Sequence 421 BP; 113 A; 100 C; 91 G; 116 T; 0 U; 1 Other;
Query Match 16.7%; Score 413.4; DB 3; Length 421;
Best Local Similarity 99.5%; Pred. No. 4.7e-59;
Matches 414; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1745 TTGCTTCTCTGCAACAGAAATGACACTGCTGGATGAGCTCAAGCTCTCAGACG 1804
DB 416 TGGCTTCTCTGCAACAGAAATGACACTGCTGGATGAGCTCAAGCTCTCAGACG 357
QY 1805 GCACTGCATGCTCTGCAACAGAAATGACACTGCTGGATGAGCTCAAGCTCTCAGACG 1864
DB 356 GCACTGCATGCTCTGCAACAGAAATGACACTGCTGGATGAGCTCAAGCTCTCAGACG 297
QY 1865 ACTGAAACTCATTTGGCTGCATATATCAGACACAGACTCACAACGCCAGAAAGCATT 1924
DB 296 ACTGAAACTCATTTGGCTGCATATATCAGACACAGACTCACAACGCCAGAAAGCATT 237
QY 1925 CTCAAGATGATGAATGTACTCAACAGAACTTCAATTAAATGCTGAAGCATTTCAGT 1984
DB 236 CTCAAGATGATGAATGTACTCAACAGAACTTCAATTAAATGCTGAAGCATTTCAGT 177
QY 1985 GCCTAGATCTGTTGTTGGCTCTTTAGAACATGATGGAGGTGAATTCCTCACTGACATG 2044
DB 176 GCCTAGATCTGTTGTTGGCTCTTTAGAACATGATGGAGGTGAATTCCTCACTGACATG 117
QY 2045 AAGTATGGACACCTTTGGTTCAGTTCAGGCGAGATTCTCCCTGTTGCTAACTGCCAGAT 2104
DB 116 AAGTATGGACACCTTTGGTTCAGTTCAGGCGAGATTCTCCCTGTTGCTAACTGCCAGAT 57
QY 2105 TTGCTTTTCAGTGTGGCTGTGGATTATCAATAGCCAGAAAGAACTCAACTGGTC 2160
DB 56 TTGCTTTTCAGTGTGGCTGTGGATTATCAATAGCCAGAAAGAACTCAACTGGTC 1

RESULT 14
AAI28605/c
ID AAI28605 standard; cDNA; 421 BP.
XX
AC AAI28605;
XX
XX 12-OCT-2001 (first entry)
DT
DE Colon tumour related determined cDNA sequence for Cosub-42.
XX
XX Human; immunotherapy; diagnosis; colon cancer; colon tumour; immunogenic;
KW gene therapy; vaccine; colonic cancer; ss.
XX
OS Homo sapiens.
XX
XX WO200149716-A2.
XX
PD 12-JUL-2001.
XX
XX 29-DEC-2000; 2000WO-US035596.
XX
PR 30-DEC-1999; 99US-00476296.
PR 10-JAN-2000; 2000US-00480321.
PR 15-FEB-2000; 2000US-00504629.
PR 06-MAR-2000; 2000US-00519444.
PR 19-MAY-2000; 2000US-00575251.
PR 29-JUN-2000; 2000US-00609448.

PR 28-AUG-2000; 2000US-00649811.
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
XX
XX WPI; 2001-441847/47.
XX
XX Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer.
XX
XX Claim 25; Page 174; 472pp; English.
XX
XX The present invention describes colon tumor associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumor associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be used
CC to treat disorders associated with decreased expression by rectifying
CC mutations or deletions in a patient's genome that affect the activity of
CC TCAPs by expressing inactive proteins or to supplement the patients own
CC production of them. Additionally, (II) may be used to produce the TCAP
CC proteins, by inserting the nucleic acids into a host cell culturing the
CC cell to express the protein. (II) and its complementary sequences may
CC also be used as DNA probes in diagnostic polymerase chain reaction (PCR)
CC and hybridisation assays to detect and quantitate the presence of similar
CC nucleic acids in samples, and therefore which patients may be in need of
CC restorative therapy. (I) may also be used as antigens in the production
CC of antibodies against TCAPs and in assays to identify modulators of TCAP
CC expression and activity. Anti-(I) antibodies and antagonists may also be
CC used to down regulate TCAP expression and activity. The anti-(I)
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of TCAPs in samples (e.g. by enzyme linked immunosorbant assay
CC (ELISA)). AA128460 to AA129512 and AA24494 to AA24523 represent
CC nucleotide and amino acid sequences given in the exemplification of the
CC present invention
XX
XX Sequence 421 BP; 113 A; 100 C; 91 G; 116 T; 0 U; 1 Other;

Query Match 16.7%; Score 413.4; DB 4; Length 421;
Best Local Similarity 99.5%; Pred. No. 4.7e-59;
Matches 414; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1745 TTGCTTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGTCTCAGTGTCTCAGCG 1804
DB 416 TGGCTTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGTCTCAGTGTCTCAGCG 357
QY 1805 GCACCTGCATGCTCTGCAGTTAACTCACTCAAGGAACCATGGCTTTGTTGGAGGTGGCTGT 1864
DB 356 GCACCTGCATGCTCTGCAGTTAACTCACTCAAGGAACCATGGCTTTGTTGGAGGTGGCTGT 297
QY 1865 ACTGAAACTCATTTGGCTGCATATATCAGACACAGACTCACAACGACCCAGAAAGCAATT 1924
DB 296 ACTGAAACTCATTTGGCTGCATATATCAGACACAGACTCACAACGACCCAGAAAGCAATT 237
QY 1925 CTCAAAGATGATGAATGTACTCAAAACAGAACTTCAATTAATTGCTGAAGCAATTTTGCAGT 1984
DB 236 CTCAAAGATGATGAATGTACTCAAAACAGAACTTCAATTAATTGCTGAAGCAATTTTGCAGT 177
QY 1985 GCCTAGAACTCTTTGCTGCTCTTTAGAACATGATGGAGGTGAATTTCTCACTGACATG 2044
DB 176 GCCTAGAACTCTTTGCTGCTCTTTAGAACATGATGGAGGTGAATTTCTCACTGACATG 117
QY 2045 AAGTATGGACACCTTTGGTTCAGTTTCAGGAGATTTCCCTGTGTGCTAACTGGCCAGAT 2104
DB 116 AAGTATGGACACCTTTGGTTCAGTTTCAGGAGATTTCCCTGTGTGCTAACTGGCCAGAT 57
QY 2105 TTGCTTTACAGTGTGGCTGTGATTAACAATAGCCAGGAAGAACTCAACTGGTC 2160
DB 56 TTGCTTTACAGTGTGGCTGTGATTAACAATAGCCAGGAAGAACTCAACTGGTC 1

RESULT 15
ABZ32791/c
ID ABZ32791 standard; cDNA; 421 BP.
XX
XX AC ABZ32791;
XX
XX DT 30-JAN-2003 (first entry)
XX
XX DE Human colon tumour related cDNA for CoSub-42 SEQ ID NO:147.
XX
XX KW Human; colon cancer; colon tumour; immunotherapy; diagnosis; cancer;
XX tumour; immune response; immunostimulant; cytostatic; vaccine; gene; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200283070-A2.
XX
XX PD 24-OCT-2002.
XX
XX PF 09-APR-2002; 2002WO-US011475.
XX
XX PR 10-APR-2001; 2001US-00833263.
XX
XX PR 03-AUG-2001; 2001US-00922217.
XX
XX PR 19-DEC-2001; 2001US-00025380.
XX
XX PA (CORI-) CORIXA CORP.
XX
XX PI Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;
PI Wang T, Jiang Y, Smith CL, King GE, Wang A, Clapper JD, Skeiky YAW;
PI Fanger GR, Vedvick TS, Carter D;
XX
XX DR WPI; 2003-067548/06.
XX
XX PT New polynucleotide, useful for the preparation of a composition for
PT stimulating an immune response against, or treating, cancer.
XX
XX PS Example 2; Page 176-177; 537pp; English.
XX
XX CC The present invention describes compounds (I) for the immunotherapy and
CC diagnosis of colon cancer. Also described: (1) a method for detecting the
CC presence of cancer in a patient; (2) a method for stimulating and/or
CC expanding T cells specific for a tumour protein; (3) an isolated T cell
CC population comprising T cells prepared by the method of (2); (4) a method
CC for stimulating an immune response in a patient; (5) a method for
CC treating cancer in a patient; and (6) a method for inhibiting the
CC development of cancer in a patient. (I) have immunostimulant and
CC cytostatic activities and can be used in vaccines. ABZ32646 to ABZ33725
CC and ABP55343 to ABP55391 represent human colon cancer/tumour related
CC sequences used in the exemplification of the present invention
XX
XX SQ Sequence 421 BP; 113 A; 100 C; 91 G; 116 T; 0 U; 1 Other;

Query Match 16.7%; Score 413.4; DB 8; Length 421;
Best Local Similarity 99.5%; Pred. No. 4.7e-59;
Matches 414; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1745 TTGCTTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGTCTCAGTGTCTCAGCG 1804
DB 416 TGGCTTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGTCTCAGTGTCTCAGCG 357
QY 1805 GCACCTGCATGCTCTGCAGTTAACTCACTCAAGGAACCATGGCTTTGTTGGAGGTGGCTGT 1864
DB 356 GCACCTGCATGCTCTGCAGTTAACTCACTCAAGGAACCATGGCTTTGTTGGAGGTGGCTGT 297
QY 1865 ACTGAAACTCATTTGGCTGCATATATCAGACACAGACTCACAACGACCCAGAAAGCAATT 1924
DB 296 ACTGAAACTCATTTGGCTGCATATATCAGACACAGACTCACAACGACCCAGAAAGCAATT 237
QY 1925 CTCAAAGATGATGAATGTACTCAAAACAGAACTTCAATTAATTGCTGAAGCAATTTTGCAGT 1984
DB 236 CTCAAAGATGATGAATGTACTCAAAACAGAACTTCAATTAATTGCTGAAGCAATTTTGCAGT 177

Qy	1985	GCCTAGAACTCTGTTGTTGGCTCTTTAGAACATGATGGAGGTGAAATTTCTCACTGACATG	2044
Db	176		
Qy	2045	AAGTATGGACACCTTTGGTCAGTTTCAGGCAGATTCTCCCTGTGTTGCTAACTGGCCAGAT	2104
Db	116		
Qy	2105	TTGCTTTTCACAGTGTGGCTGTGGATTATACAATAGCCAGGAAGAACTCAACTGGTC	2160
Db	56		

Search completed: September 6, 2005, 12:00:12
Job time : 1323 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2005, 04:44:12 ; Search time 178 Seconds
(without alignments)
1639.804 Million cell updates/sec

Title: US-10-616-263-30

Perfect score: 2957

Sequence: 1 MSRLAKPKSLCKSEPLTTE.....VAVETANLIWLSYVIEDKN 570

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Uniprot_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2936	99.3	570	1 MKKS HUMAN	Q9npj1 homo sapien
2	2919	98.7	570	2 Q81222	Q81222 homo sapien
3	2904	98.2	570	2 Q9H6M6	Q9H6M6 homo sapien
4	2334	78.9	570	2 Q8BQ33	Q8BQ33 m mus-muscu
5	2331	78.8	570	1 MKKS MOUSE	Q9j170 mus musculu
6	2021	68.3	502	2 Q9DBF3	Q9dbf3 mus musculu
7	1185.5	40.1	563	2 Q7ZVV0	Q7zvv0 brachydanio
8	516	17.5	121	2 Q8K3C1	Q8k3c1 mus musculu
9	279	9.4	553	2 Q8ZTF8	Q8ztf8 pyrobaculum
10	270	9.1	557	2 Q59662	Q59662 pyrodictium
11	259.5	8.7	563	2 Q877H0	Q877h0 methanosarc
12	257.5	8.7	538	1 THSB METTH	Q26885 methanobact
13	256.5	8.7	560	1 THSA SULSH	P4619 sulfolobus
14	255	8.6	490	1 THSA SULAC	Q9v2t5 sulfolobus
15	253.5	8.6	546	2 Q6L132	Q6l132 picropilus
16	253.5	8.6	559	1 THSA SULSO	Q9v289 sulfolobus
17	253	8.6	552	2 Q8TU11	Q8tu11 methanosarc
18	252.5	8.5	559	1 THSA SULTO	Q24734 sulfolobus
19	248.5	8.4	554	1 THSA AERPE	Q9ydk6 aeropyrum p
20	246	8.3	551	2 Q8PX43	Q8px43 methanosarc
21	244	8.3	549	2 Q8ZVV0	Q8zvv0 pyrobaculum
22	243	8.2	542	2 Q7RR86	Q7rr86 pyrobaculum
23	238	8.0	545	1 THSA ARCFU	Q28045 archaeoglob
24	236	8.0	554	2 Q6BY12	Q6by12 debaryomyce
25	235	7.9	542	1 TCPO SCHPO	Q8pxx0 methanosarc
26	234.5	7.9	546	1 TCPO SCHPO	P78921 schizosach
27	234.5	7.9	561	2 Q6C9H3	Q6c9h3 yarrowia li
28	234.5	7.9	572	2 Q59663	Q59663 pyrodictium
29	234	7.9	495	1 THSB SULAC	Q9v2t4 sulfolobus
30	233	7.9	528	2 Q8IDC7	Q8idc7 plasmodium
31	232	7.8	517	2 Q7RK64	Q7rk64 plasmodium

32 232 7.8 535 1 THSG SULSH
33 230 7.8 543 2 Q877G8
34 229.5 7.8 542 1 THS METJA
35 229 7.7 542 2 Q815C4
36 228.5 7.7 502 1 THSB DESMO
37 227.5 7.7 559 2 Q6CY04
38 227 7.7 543 2 Q8THU8
39 226 7.6 535 1 THSG SULSO
40 226 7.6 548 1 THSB AERPE
41 226 7.6 558 2 Q751B1
42 225.5 7.6 545 1 THSA THEAC
43 225 7.6 544 1 TCFG HUMAN
44 225 7.6 563 2 Q6CRG0
45 224.5 7.6 547 2 Q751Y6

ALIGNMENTS

RESULT 1
MKKS HUMAN
ID MKKS_HUMAN STANDARD; PRT; 570 AA.
AC Q9NFU1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE McKusick-Kaufman/Bardet-Biedl syndromes putative chaperonin.
GN Name=MKKS;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A., VARIANTS MKKS CYS-37; TYR-84 AND SER-242, AND
RP VARIANTS VAL-49 AND CYS-517.
RX MEDLINE=20264373; PubMed=10802561; DOI=10.1038/75637;
RA Stone D.L., Slavotinek A.M., Bouffard G.G., Banerjee-Basu S.,
RA Baxevasis A.D., Barr M., Biesecker L.G.;
RT "Mutation of a gene encoding a putative chaperonin causes McKusick-
RT Kaufman syndrome.";
RN Nat. Genet. 25:79-82 (2000).

SEQUENCE FROM N.A.
RA Deloukas P., Matthews L.H., Ashurst J.L., Burton J., Gilbert J.G.R.,
RA Bailey J., Barlow K.P., Bates K.N., Bead L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.V., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leharvalho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Rose M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
RA Skuce C.D., Smith M.U., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromamore A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,

Q9hh21 sulfolobus
Q877g8 methanococ
Q58405 methanococ
Q815c4 plasmodium
Q9v2t3 desulfuroco
Q6cy04 kluyveromyc
Q8thu8 methanosarc
Q9v2t7 sulfolobus
Q9va66 aeropyrum p
Q751b1 ashbya goss
P48424 thermoplasm
P49368 homo sapien
Q6crg0 kluyveromyc
Q751y6 neurospora


```
Db 1 MSRLKPKSLCKSEPTTTRVRLTSLVKRIIVTSCYGPSGRKLQHLNFGGYYVCTTSQS 60
Qy 61 SALLSHLLVTHPIKILITASTASIONHVSSFDGCLFTAILCCNLNIENVRGLTPTTVIRLN 120
Db 61 SALLSHLLVTHPIKILITASTASIONHVSSFDGCLFTAILCCNLNIENVRGLTPTTVIRLN 120
Qy 121 KHLLSLCISYLSKSDTCGRIPVDFSSQTQLCLVRSILTSKPACMLTRKETEHSALILR 180
Db 121 KHLLSLCISYLSKSDTCGRIPVDFSSQTQLCLVRSILTSKPACMLTRKETEHSALILR 180
Qy 181 AFLTTPENAEGHIIILGKSLIPLKQGVDSITVPLGILTEMSEVQMLRLLPIKKSALK 240
Db 181 AFLTTPENAEGHIIILGKSLIPLKQGVDSITVPLGILTEMSEVQMLRLLPIKKSALK 240
Qy 241 VALFCTTSLGSDTSITGEGTVVSVGSLENAVLQDLNLRQLISDHVDLVLCKVHPS 300
Db 241 VALFCTTSLGSDTSITGEGTVVSVGSLENAVLQDLNLRQLISDHVDLVLCKVHPS 300
Qy 301 LKQFLNHRHIIAIDRIGVTLMPELTKWTGTQPIGSLGSI CPNSYSGVSKDVCTAKFGSQHF 360
Db 301 LKQFLNHRHIIAIDRIGVTLMPELTKWTGTQPIGSLGSI CPNSYSGVSKDVCTAKFGSKHF 360
Qy 361 PHILPNEATICSLLCNRNNDTAWDELKLTCTQALHVLQTLKEPVALLGCGCTETHLAAY 420
Db 361 PHILPNEATICSLLCNRNNDTAWDELKLTCTQALHVLQTLKEPVALLGCGCTETHLAAY 420
Qy 421 IRKTHNDPESILKDDCTQTELQIAEAFCSALESVVGSLEHDDGGGILLTDMKYGHLSV 480
Db 421 IRKTHNDPESILKDDCTQTELQIAEAFCSALESVVGSLEHDDGGGILLTDMKYGHLSV 480
Qy 481 QADSPCVANWPDLLSCQCGLYNSQBELNWSFLRSTRPFPVQSCLPHEAVGSASNLTLTD 540
Db 481 QADSPCVANWPDLLSCQCGLYNSQBELNWSFLRSTRPFPVQSCLPHEAVGSASNLTLTD 540
Qy 541 CLTAKLSGLQVAVETANLILDLISYVIEDKN 570
Db 541 CLTAKLSGLQVAVETANLILDLISYVIEDKN 570

RESULT 2
Q81222 ID Q81222 PRELIMINARY; PRT; 570 AA.
AC Q81222;
DC 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE McKusick-Kaufman syndrome protein.
GN Name=MKKS;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=22398257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uadin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gumaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalius D.E., Schnurch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
```

```
RT *Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC028973; AAH28973.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0051082; F:unfolded protein binding; IEA.
DR GO; GO:0006457; P:protein folding; IEA.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; Cpn60 TCP1; 1.
KW ATP-binding; Chaperone.
SQ SEQUENCE 570 AA; 62330 MW; 01E57FF8AF7BA80 CRC64;

Query Match 98.7%; Score 2919; DB 2; Length 570;
Best Local Similarity 99.1%; Pred. No. 7.3e-216;
Matches 565; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MSRLKPKSLCKSEPTTTRVRLTSLVKRIIVTSCYGPSGRKLQHLNFGGYYVCTTSQS 60
Db 1 MSRLKPKSLCKSEPTTTRVRLTSLVKRIIVTSCYGPSGRKLQHLNFGGYYVCTTSQS 60
Qy 61 SALLSHLLVTHPIKILITASTASIONHVSSFDGCLFTAILCCNLNIENVRGLTPTTVIRLN 120
Db 61 SALLSHLLVTHPIKILITASTASIONHVSSFDGCLFTAILCCNLNIENVRGLTPTTVIRLN 120
Qy 121 KHLLSLCISYLSKSDTCGRIPVDFSSQTQLCLVRSILTSKPACMLTRKETEHSALILR 180
Db 121 KHLLSLCISYLSKSDTCGRIPVDFSSQTQLCLVRSILTSKPACMLTRKETEHSALILR 180
Qy 181 AFLTTPENAEGHIIILGKSLIPLKQGVDSITVPLGILTEMSEVQMLRLLPIKKSALK 240
Db 181 AFLTTPENAEGHIIILGKSLIPLKQGVDSITVPLGILTEMSEVQMLRLLPIKKSALK 240
Qy 241 VALFCTTSLGSDTSITGEGTVVSVGSLENAVLQDLNLRQLISDHVDLVLCKVHPS 300
Db 241 VALFCTTSLGSDTSITGEGTVVSVGSLENAVLQDLNLRQLISDHVDLVLCKVHPS 300
Qy 301 LKQFLNHRHIIAIDRIGVTLMPELTKWTGTQPIGSLGSI CPNSYSGVSKDVCTAKFGSQHF 360
Db 301 LKQFLNHRHIIAIDRIGVTLMPELTKWTGTQPIGSLGSI CPNSYSGVSKDVCTAKFGSKHF 360
Qy 361 PHILPNEATICSLLCNRNNDTAWDELKLTCTQALHVLQTLKEPVALLGCGCTETHLAAY 420
Db 361 PHILPNEATICSLLCNRNNDTAWDELKLTCTQALHVLQTLKEPVALLGCGCTETHLAAY 420
Qy 421 IRKTHNDPESILKDDCTQTELQIAEAFCSALESVVGSLEHDDGGGILLTDMKYGHLSV 480
Db 421 IRKTHNDPESILKDDCTQTELQIAEAFCSALESVVGSLEHDDGGGILLTDMKYGHLSV 480
Qy 481 QADSPCVANWPDLLSCQCGLYNSQBELNWSFLRSTRPFPVQSCLPHEAVGSASNLTLTD 540
Db 481 QADSPCVANWPDLLSCQCGLYNSQBELNWSFLRSTRPFPVQSCLPHEAVGSASNLTLTD 540
Qy 541 CLTAKLSGLQVAVETANLILDLISYVIEDKN 570
Db 541 CLTAKLSGLQVAVETANLILDLISYVIEDKN 570

RESULT 3
Q9H6M6 ID Q9H6M6 PRELIMINARY; PRT; 570 AA.
AC Q9H6M6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein FLJ22088.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA	Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA	Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AK025741; BAB15230.1; -
DR	GO; GO:0005524; F:ATP binding; IEA.
DR	GO; GO:0005524; F:unfolding protein binding; IEA.
DR	GO; GO:0005524; F:protein folding; IEA.
DR	InterPro; IPR002423; Cpn60/TCP-1.
DR	InterPro; IPR008950; GroEL-ATPase.
DR	Pfam; PF00118; Cpn60_TCP1; 1.
KW	ATP-binding; Chaperone.
SQ	SEQUENCE 570 AA; 62347 MW; 2F3FE9459F472344 CRC64;
Query Match 98.2%; Score 2904; DB 2; Length 570;	
Best Local Similarity 98.8%; Pred. No. 1e-214; 6; Indels 0; Gaps 0;	
Matches 562; Conservative 2; Mismatches	
Qy	1 MSRLKAKPSLCKSEPLTTERVTTLSVLKRIIVTSCYGPSGLKQLHNGFGGVCTTSQS 60
Db	1 MSRLKAKPSLCKSEPLTTERVTTLSVLKRIIVTSCYGPSGLKQLHNGFGGVCTTSQS 60
Qy	61 SALLSHLLVTHPIKILITASIQNHSVSDCGLFTAILCNLIENVQRLGTPPTVIRLN 120
Db	61 SALLSHLLVTHPIKILITASIQNHSVSDCGLFTAILCNLIENVQRLGTPPTVIRLN 120
Qy	121 KHLISLCISVLKSDTCGRIPDVSFTOILLCLVRSILTSKPACMLTRKETEHVSALIIR 180
Db	121 KHLISLCISVLKSDTCGRIPDVSFTOILLCLVRSILTSKPACMLTRKETEHVSALIIR 180
Qy	181 AFLTTPENAEGHILGKSLIPLKQORVIDSTVLPGLIEMSEVQLMRLLPIKKSALK 240
Db	181 AFLTTPENAEGHILGKSLIPLKQORVIDSTVLPGLIEMSEVQLMRLLPIKKSALK 240
Qy	241 VALFCTTSLGSDTSDEGTGVVSYGVSLENAVLDQLNLGRQLISDHVDLVLCQKVIHPS 300
Db	241 VALFCTTSLGSDTSDEGTGVVSYGVSLENAVLDQLNLGRQLISDHVDLVLCQKVIHPS 300
Qy	301 LKQFLNHRHIIADRIQVTLMEPLTKMTGTPQIPGSLGSLCPNSYSGVKDCTAKFGSHF 360
Db	301 LKQFLNHRHIIADRIQVTLMEPLTKMTGTPQIPGSLGSLCPNSYSGVKDCTAKFGSKHF 360
Qy	361 FLHIPNEATCSLLLCNRNDTANDELKLTCTQALHVLQTLKEPFWALLGGGCTETHLAAY 420
Db	361 FLHIPNEATCSLLLCNRNDTANDELKLTCTQALHVLQTLKEPFWALLGGGCTETHLAAY 420
Qy	421 IRHKTHNDPSILKDDCTQELQIAEAPCSALESVSGSLEHGDGEIITDMKYGHLWSV 480
Db	421 IRHKTHNDPSILKDDCTQELQIAEAPCSALESVSGSLEHGDGEIITDMKYGHLWSV 480
Qy	481 QADSPCVANWPDLLSCGGLYNSQBELNWSFLRTPRPVPOCSLPHEAVGSASNLITLD 540
Db	481 HADSPCVANWPDLLSCGGLYNSQBELNWSFLRTPRPVPOCSLPHEAVGSASNLITLD 540
Qy	541 CLTKLISGLQVAVETANLIWDLGVYIEDKN 570
Db	541 CLTKLISGLQVAVETANLIWDLGVYIEDKN 570
RESULT 4	
ID	Q8BGQ3 PRELIMINARY; PRT; 570 AA.
AC	Q8BGQ3;
DT	01-MAR-2003 (TrEMBLrel. 23, Created)
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT	25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE	Mus musculus adult male olfactory brain cDNA, RIKEN full-length
DE	enriched library, clone:6430582D06 product:McKusick-Kaufman syndrome
DE	protein, full insert sequence (Mus musculus adult male olfactory brain

DE	CDNA, RIKEN full-length enriched library, clone:6430593I17
DE	Product:McKusick-Kaufman syndrome protein, full insert sequence).
GN	Name=Mkks;
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX	MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA	Carninci P., Hayashizaki Y.;
RA	"High-efficiency full-length cDNA cloning.";
RT	Meth. Enzymol. 303:19-44 (1999).
RL	Nature 409:685-690 (2001).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX	MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA	RIKEN FANTOM Consortium;
RA	"Functional annotation of a full-length mouse cDNA collection.";
RT	Nature 409:685-690 (2001).
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RA	The FANTOM Consortium;
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;
RT	"Analysis of the mouse transcriptome based on functional annotation of
RT	60,770 full-length cDNAs.";
RT	Nature 420:563-573 (2002).
RN	[4]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX	MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA	Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA	Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA	Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA	Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA	Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA	Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA	Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT	"RIKEN integrated sequence analysis (RISA) system-384-format
RT	sequencing pipeline with 384 multicapillary sequencer.";
RT	Genome Res. 10:1757-1771 (2000).
RN	[5]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX	MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA	Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA	Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA	Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA	Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA	Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA	Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA	Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT	"RIKEN integrated sequence analysis (RISA) system-384-format
RT	sequencing pipeline with 384 multicapillary sequencer.";
RT	Genome Res. 10:1757-1771 (2000).
RN	[6]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RA	Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA	Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA	Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA	Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA	Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA	Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA	Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA	Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA	Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA	Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,
RA	Tomaru A., Taya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RL	EMBL; AK032528; BAC27912.1; -
DR	EMBL; AK032528; BAC27912.1; -
DR	MGI; MGI:1891836; Mkks.
DR	GO; GO:0005524; F:ATP binding; IEA.
DR	GO; GO:0051082; F:unfolding protein binding; IEA.
DR	GO; GO:0006457; P:protein folding; IEA.

```
DR InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; Cpn60 TCP1; 1.
KW ATP-binding; Chaperone.
SQ SEQUENCE 570 AA; 61923 MW; 7E2AA28D43028800 CRC64;

Query Match 78.9%; Score 2334; DB 2; Length 570;
Best Local Similarity 76.7%; Pred. No. 7.4e-171;
Matches 437; Conservative 62; Mismatches 71; Indels 0; Gaps 0;

QY 1 MSRLKAKPSICKSEPLTTRVRLTSLVKRIVTCVGPGRGLKQLHNGFGGVCTTSQS 60
DB 1 MSRLKAKPSICKSEPLTTRVRLTSLVKRIVTCVGPGRGLKQLHNGFGGVCTTSQS 60

QY 61 SALLSHLLVTHPIKLTASTQNHVSSFCGFLFTAILCCNLLENVQRLGLTPTTVIRLN 120
DB 61 SALLSHLLVTHPIKLTASTQNHVSSFCGFLFTAILCCNLLENVQRLGLTPTTVIRLN 120

QY 121 KHLISLCISYLSKSDTCGRIPVDPSSTQIILLCLVRSILTSKPACMLTRKETEHSALILR 180
DB 121 KHLISLCISYLSKSDTCGRIPVDPSSTQIILLCLVRSILTSKPACMLTRKETEHSALILR 180

QY 121 KYLLSLCTSYLSKSEACSCRIIPVDFRSTHTFLSLVHSILTSKPACMLTRKETHIGALILK 180
DB 121 KYLLSLCTSYLSKSEACSCRIIPVDFRSTHTFLSLVHSILTSKPACMLTRKETHIGALILK 180

QY 181 AFLTTPENAEHIIIGKSLIIVPLKQGVDSVPLGILIESEVQLMRLLPIKKTALK 240
DB 181 AFLTTPENAEHIIIGKSLIIVPLKQGVDSVPLGILIESEVQLMRLLPIKKTALK 240

QY 181 AFLTTPESTERVMVGLKSIIVPLKQGVDSVPLGILIESEVQLMRLLPIKKTALK 240
DB 181 AFLTTPESTERVMVGLKSIIVPLKQGVDSVPLGILIESEVQLMRLLPIKKTALK 240

QY 241 VALFCTTSLGSDTSFGTGVVSVYGVLENVLDQLNLGRQLISDHVDLVLCQKVIHPS 300
DB 241 VALFCTTSLGSDTSFGTGVVSVYGVLENVLDQLNLGRQLISDHVDLVLCQKVIHPS 300

QY 301 LKQFLNMHRIIADRIQVTLMEPLTKMTGTQPIGSLGSCIPNSYGVSKDVCTAKFGSQHF 360
DB 301 LKQFLNMHRIIADRIQVTLMEPLTKMTGTQPIGSLGSCIPNSYGVSKDVCTAKFGSQHF 360

QY 361 FHLIPNEATVCTLLCSNRNDTAWDELKLTCTQTAHVLQTLKEPWWLLGGGCTETHLAAY 420
DB 361 FHLIPNEATVCTLLCSNRNDTAWDELKLTCTQTAHVLQTLKEPWWLLGGGCTETHLAAY 420

QY 421 IRKTHNDPESILKDDCTQTELOLIAEAFCSALESVVGSLEHDGGEIITDMKYGHLWSV 480
DB 421 IRKTHNDPESILKDDCTQTELOLIAEAFCSALESVVGSLEHDGGEIITDMKYGHLWSV 480

QY 481 QADSPCVANWPDLLSQCGGLYNSQBELNWSFLRSTRPFVPSQCLPHEAVGASNLITLD 540
DB 481 QADSPCVANWPDLLSQCGGLYNSQBELNWSFLRSTRPFVPSQCLPHEAVGASNLITLD 540

QY 541 CLTAKLSGLQVAVETANLWLSVYIEDKN 570
DB 541 CLTAKLSGLQVAVETANLWLSVYIEDKN 570

RESULT 5
MKKS_MOUSE STANDARD; PRT; 570 AA.
AC Q9J170;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE McKusick-Kaufman/Bardet-Biedl syndromes putative chaperonin.
GN Name=Mkke;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20264373; PubMed=10802661; DOI=10.1038/75637;
RA Stone D.L., Slavotinek A., Bouffard G.G., Banerjee-Basu S.,
RA Baxevanis A.D., Barr M., Biesecker L.G.;
RT "Mutation of a gene encoding a putative chaperonin causes McKusick-
RT Kaufman syndrome."
RL Nat. Genet. 25:79-82(2000).
CC -!- FUNCTION: May play a role in protein processing in limb, cardiac
```

```
CC and reproductive system development.
CC -!- TISSUE SPECIFICITY: Widely expressed in adult and fetal tissues.
CC -!- SIMILARITY: Belongs to the TCP-1 chaperonin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF254074; AAF73965.1; -.
CC MGD; MGI:1891836; Mkke.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC InterPro; IPR008950; GroEL-ATPase.
CC Pfam; PF00118; Cpn60 TCP1; 1.
KW ATP-binding; Chaperone.
FT NP_BIND 192 199 ATP (Potential).
SQ SEQUENCE 570 AA; 61951 MW; 740CDB8D44AE0100 CRC64;

Query Match 78.8%; Score 2331; DB 1; Length 570;
Best Local Similarity 76.5%; Pred. No. 1.3e-170;
Matches 436; Conservative 63; Mismatches 71; Indels 0; Gaps 0;

QY 1 MSRLKAKPSICKSEPLTTRVRLTSLVKRIVTCVGPGRGLKQLHNGFGGVCTTSQS 60
DB 1 MSRLKAKPSICKSEPLTTRVRLTSLVKRIVTCVGPGRGLKQLHNGFGGVCTTSQS 60

QY 61 SALLSHLLVTHPIKLTASTQNHVSSFCGFLFTAILCCNLLENVQRLGLTPTTVIRLN 120
DB 61 SALLSHLLVTHPIKLTASTQNHVSSFCGFLFTAILCCNLLENVQRLGLTPTTVIRLN 120

QY 121 KHLISLCISYLSKSDTCGRIPVDPSSTQIILLCLVRSILTSKPACMLTRKETEHSALILR 180
DB 121 KHLISLCISYLSKSDTCGRIPVDPSSTQIILLCLVRSILTSKPACMLTRKETEHSALILR 180

QY 121 KYLLSLCTSYLSKSEACSCRIIPVDFRSTHTFLSLVHSILTSKPACMLTRKETHIGALILK 180
DB 121 KYLLSLCTSYLSKSEACSCRIIPVDFRSTHTFLSLVHSILTSKPACMLTRKETHIGALILK 180

QY 181 AFLTTPENAEHIIIGKSLIIVPLKQGVDSVPLGILIESEVQLMRLLPIKKTALK 240
DB 181 AFLTTPENAEHIIIGKSLIIVPLKQGVDSVPLGILIESEVQLMRLLPIKKTALK 240

QY 241 VALFCTTSLGSDTSFGTGVVSVYGVLENVLDQLNLGRQLISDHVDLVLCQKVIHPS 300
DB 241 VALFCTTSLGSDTSFGTGVVSVYGVLENVLDQLNLGRQLISDHVDLVLCQKVIHPS 300

QY 301 LKQFLNMHRIIADRIQVTLMEPLTKMTGTQPIGSLGSCIPNSYGVSKDVCTAKFGSQHF 360
DB 301 LKQFLNMHRIIADRIQVTLMEPLTKMTGTQPIGSLGSCIPNSYGVSKDVCTAKFGSQHF 360

QY 361 FHLIPNEATVCTLLCSNRNDTAWDELKLTCTQTAHVLQTLKEPWWLLGGGCTETHLAAY 420
DB 361 FHLIPNEATVCTLLCSNRNDTAWDELKLTCTQTAHVLQTLKEPWWLLGGGCTETHLAAY 420

QY 421 IRKTHNDPESILKDDCTQTELOLIAEAFCSALESVVGSLEHDGGEIITDMKYGHLWSV 480
DB 421 IRKTHNDPESILKDDCTQTELOLIAEAFCSALESVVGSLEHDGGEIITDMKYGHLWSV 480

QY 481 QADSPCVANWPDLLSQCGGLYNSQBELNWSFLRSTRPFVPSQCLPHEAVGASNLITLD 540
DB 481 QADSPCVANWPDLLSQCGGLYNSQBELNWSFLRSTRPFVPSQCLPHEAVGASNLITLD 540

QY 541 CLTAKLSGLQVAVETANLWLSVYIEDKN 570
DB 541 CLTAKLSGLQVAVETANLWLSVYIEDKN 570

RESULT 6
Q9DBF3 PRELIMINARY; PRT; 502 AA.
ID Q9DBF3
AC Q9DBF3;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
```

DE Mus musculus adult male liver cDNA, RIKEN full-length enriched
 DE library, clone:1300013E18 product:McKusick-Kaufman syndrome protein,
 DE full insert sequence.
 GN Name=Mkks;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Liver;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "high-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Liver;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Liver;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Liver;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Liver;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Liver;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Liver;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).

DR InterPro; IPR002423; Cpn60/TCP-1.
 DR InterPro; IPR008950; GroEL-ATPase.
 DR Pfam; PF00118; Cpn60_TCP1; 1.
 KW ATP-binding; Chaperone.
 SQ SEQUENCE 502 AA; 54611 MW; 6C4407F4BF19A6AD CRC64;
 Query Match 68.3%; Score 2021; DB 2; Length 502;
 Best Local Similarity 67.5%; Pred. No. 7.5e-147;
 Matches 385; Conservative 53; Mismatches 64; Indels 68; Gaps 1;
 QY 1 MSRLAKPSLCKSEKPLTTERVTTLSVLKRVITVTCYGPSGRGLKQLHNGFGYVCTTSOS 60
 DB 1 MSRLAKPSLCKSEKPLTTERVTTLSVLKRVITVTCYGPSGRGLKQLHNGFGYVCTTSOS 60
 QY 61 SALLSHLVTHPIKILITASIIONHVSFSDGCLFTAILCCNLIENVRQLGLTPPTVIRLN 120
 DB 61 SALLSHLVTHPIKILITASIIONHVSFSDGCLFTAILCCNLIENVRQLGLTPPTVIRLN 120
 QY 121 KHLISICISYLSKSDTCGCRIPVDFFSTQILLGLVRSILTSKPCMLTRKETEHVSALILR 180
 DB 121 KHLISICISYLSKSDTCGCRIPVDFFSTQILLGLVRSILTSKPCMLTRKETEHVSALILR 180
 QY 121 KYLLSLCTSYLSKSEACSCRIIPVDFRSTHTFSLVHSILTSKPCMLTRKETDHIGALILK 180
 DB 121 KYLLSLCTSYLSKSEACSCRIIPVDFRSTHTFSLVHSILTSKPCMLTRKETDHIGALILK 180
 QY 181 AFLLTIPENAEGHIIILGKSLIVPLRGQRVIDSTVLPGLIEMSEVQLMELLPIKKSTALK 240
 DB 181 AFLLTIPENAEGHIIILGKSLIVPLRGQRVIDSTVLPGLIEMSEVQLMELLPIKKSTALK 240
 QY 181 AFLLTIPESTEERMWLGKSIIVPLKGQRTDSTVLPGLIEASEVQLRELLPTQK---- 235
 DB 181 AFLLTIPESTEERMWLGKSIIVPLKGQRTDSTVLPGLIEASEVQLRELLPTQK---- 235
 QY 241 VALFCTTSLSGDSTGEGTVVSYGVLENVLDQLNLGRQLISDHVDLVLCQKVIHPS 300
 DB 241 VALFCTTSLSGDSTGEGTVVSYGVLENVLDQLNLGRQLISDHVDLVLCQKVIHPS 300
 QY 236 ----- 235
 DB 236 ----- 235
 QY 301 LKQFLNMHRIIALDRIGVTLMETKMTGTQPIGSIIGSPNSYGVSKDVCTAKFGSOHP 360
 DB 301 LKQFLNMHRIIALDRIGVTLMETKMTGTQPIGSIIGSPNSYGVSKDVCTAKFGSOHP 360
 QY 236 ----FFSERHVMALDRVGVTLMSLSKVGTATPIGSLNPVSTYGVSKDVCSARFGSKHP 292
 DB 236 ----FFSERHVMALDRVGVTLMSLSKVGTATPIGSLNPVSTYGVSKDVCSARFGSKHP 292
 QY 361 FHLIPNEATICSILLNCNRNDTAWDELKLTCTQALHVLQTLKEPWALLGGCGTETHLAAY 420
 DB 293 FHLIPNEATVCTLLCSNRNDTAWDELKLTCTQAMHVLQTLKEPWALLGGCGTETHLAAY 352
 QY 421 IRKHTNDPESILKODECTQTELOLIAEAFCSALESVVGSLHEDGGEILTMKYGHLSV 480
 DB 353 VRHKVHHEAEAIVRDDGCTQAKLHVAAEAFCSALESVAGSLHEDGGEILIDTKYGHLSV 412
 QY 481 QADSPCVANWPDLLSCGGLYNSQBELNWSFLRSTRFPVPSQCLPHEAVGSASNLTD 540
 DB 413 QADSAVGVNWDTLSCGGLYNSQBELNWSFLRSTRFPVPSQCLPHEAVGSASNLTD 472
 QY 541 CUTAKLSGLQAVETANLIDLSYVIEDKN 570
 DB 473 CFTAKLSGLQAVETANLIDLSYVIEDKN 502
 [1]
 RESULT 7
 Q7ZVV0 PRELIMINARY; PRT; 563 AA.
 ID Q7ZVV0
 AC Q7ZVV0; PRELIMINARY; PRT; 563 AA.
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Similar to McKusick-Kaufman syndrome.
 GN ORFNames=zgc:55608;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 ON NCBI_TaxID=7955;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AB; TISSUE=Whole body;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Faney J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.C., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywiński M.I., Skalska U., Smalhus D.E., Scherch A., Schein J.E., Jones S.J., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human RT and mouse cDNA sequences."; Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	SRAIN=AB; TISSUE=whole body;
RA	Strauberg R.;
RL	Submitted (JAN-2003) to the EMBL/GenBank/DBSJ databases.
DR	EMBL; BC045401; AAH45401.1; -.
DR	ZFIN; ZDB-GENE-040426-757; zgc:55608.
DR	GO; GO:0005524; F:ATP binding; IEA.
DR	GO; GO:0051082; F:unfolded protein binding; IEA.
DR	GO; GO:0006457; P:protein folding; IEA.
DR	InterPro; IPR002423; Cpn60/TCP-1.
DR	Pfam; PF00118; Cpn60_TCP1; 1.
KW	ATP-binding; Chaperone.
SQ	SEQUENCE 563 AA; 61538 MW; B94AS4F942483796 CRC64;
	Query Match 40.1%; Score 1185.5; DB 2; Length 563;
	Best Local Similarity 41.1%; Pred. No. 1.7e-82;
	Matches 237; Conservative 123; Mismatches 196; Indels 21; Gaps 5
Qy	1 MSRLAEKKPSLCKSEPLTTERVRTLVLKRVITSCYGPSGLKQLHNGFGYGVCTTSQS 60
Db	1 MSRTSKKKPALCCTDEPLNSNTICQKITLRNLSTAYGFTGRLLQIHNVGGHVLTSTS 60
Qy	61 SALLSHLVTHPIKLITASTONHVSSFDCLFTAILCCNLNIENVORGLTPTTVILN 120
Db	61 TALDKRLEMSPELKLIALTAQHHTTRYSDSLGFMGIFTLTIENIKYGLRTSAIKVY 120
Qy	121 KHLISLCITSYLKSDTCGRIPDVFDSSTQILLCLVRSILTSPACMLTRKETHEVSALIR 180
Db	121 KHLVEQCNYLKGSDCGCKVPFESSCDLSVALARSMTSPACMLDSREMQQTSSLIQ 180
Qy	181 AFLTIPENAEHHIILKSLIVLPKQRVIDSTVLPGLIEMSEVOLMRLLPIPKSTALK 240
Db	181 AFLYSIPCNSSGTAGCFRTVTIGIEGQSVNHSSVFPGLLDVPEMLPGDLERLGDPFK 240
Qy	241 VALFCITLSDGSTGTBEGTVVSVGSLENAVLQOLLNGLRLISDHVDLVLCVKVIHPS 300
Db	241 VWLFVGVSLSGDISEVGDVALEVHRLGNPERDLLOQLLKIGEQA VKDKVSLFACQKVHPV 300
Qy	301 LKQFLNMHRITAI DRIGVTLMELPTTKMTGTGPISGLSGICEP- NSYGSKVDVCTAKFGSQH 359
Db	301 LOHYLRHEVVVIERGLALMEPAKITGARAVASLSLVPMWAYGLVAGLCFQDCGSKK 360
Qy	360 FFHLIPNE-ATICSLILCNRNDDTAWEDELKLTCTOTALHVLTQLKEPWALLGGGCTETHLA 418
Db	361 LLQLLSKHAIAISTMWLCHRNETMLELKMTCQRAEHVLRITLREPALLYALLGGGCTETQLA 420
Qy	419 AYIRHKTNPDPESTLKDECTQTELQLIAEAFCSALESVVGSLHDGGEIITDMKYGHILW 478
Db	421 THISHMNQSTAPTTAAALGISHSFEFLMAVESFRSLLAVALSLSDHGDDOCDILDTYGRHW 480
Qy	479 SVQADSPCANWPDLLSCQGGLYNSEBELNWSFLRSTRFPVP-----QSCLPHEAVGS 533
Db	481 VMD-----ISSQTEVKHTCCGGLLEDNRSLNBEKHTLNTACQTFSPVFNSENSTQPK----- 530
Qy	534 ASNLTDCLTAKLSGLQAVETANLIWLDSVIEDKN 570

Db 21 EALRNILAAKVAELVLSKSLGPRGLDKMLVDSFGD-VTTNDGATILKEMEIQHPAAKL 79
 Qy 77 LTASIQNHVSFSCGLFTAILCCNLJ-ENVRGLGPTTVIRLNKHLISICISYLK 132
 Db 80 MVEVAKAQDAEVDG-GTTSVAVLAGMLLDRAENLDENIHTTIIIEGKKALDYALGELS 138
 Qy 133 SDTCGRIPVDFSTQILLCLVRSILTSK----PACMLTRKETEHVSALILRAPLTIPE 188
 Db 139 K----IGVKVDNDRLKRIASTSLYSKYVSGATW-----DKLTDMAVEAVLRVAEP 198
 Qy 189 NAEG--HIIIGKSLIVPLKQGVDSVTLPGILIEHSEVQMLRLLPIKKSTALKVALFCT 246
 Db 189 RGDGTYSVRLDRVKIEKKGSLDSQVGVILDKVWH--PGMPKRVENAVIVLIDAP 246
 Qy 247 TISGDTSDTCGTGTVVSGVSLN---AVLDLNLGRLIS-----DHVD-- 299
 Db 247 L-----EVKPEITAKINITSPDQIKAFIDEEAKLRWDRIYEIAVERMKRDMGMEP 300
 Qy 290 ----LVLCKQVHPSLQKFLNMRHIIAIDRIGVTLMPLTKMTGTQPIGSLGICPNSYG 345
 Db 301 KAGIVITQKIDVQAQHLAKKGMVAVRRVGRSDLEKLEVATGGRIVSSLRDLKPEDLG 360
 Qy 346 SVKDVCTAKGSOH--FFHLIPNEATICSILLCNENDTAWDELKLTCTQTAHVLQTLKE 403
 Db 361 FAKLVEERKVGNDKQWFIIEGCPNPKAV-TILLRGANDMVLDEAERSLNDALHVLRLNRK 419
 Qy 404 PWALLGGCTETHLAAVIRKHTNDPESILKDECTQTELQIAEAFCSALESVVGSLH 463
 Db 420 PLIVPGGAVEVELMRLR-----KYAESLGGKEQLAVEAYADALEEIPMVLA 468
 Qy 464 DGG-----EILTDMKYGH 476
 Db 469 SAGMDALQALMDLRLH 485

RESULT 11

Q877H0 PRELIMINARY; PRT; 563 AA.
 AC Q877H0;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Chaperonin alpha subunit.
 OS Acidianus tengchongensis.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Acidianus.
 OX NCBI_TaxID=146920;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=S5;
 RL Ma Q., Zhang Y.;
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY254173; AAP04526.1; -;
 DR HSSP; P48424; IASS.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0051082; F:unfolded protein binding; IEA.
 DR GO; GO:0006457; P:protein folding; IEA.
 DR InterPro; IPR002194; Chaperonin_TCP-1.
 DR InterPro; IPR001844; Chaperin_Cpn60.
 DR InterPro; IPR002423; Cpn60/TCP-1.
 DR InterPro; IPR008950; GroEL-ATPase.
 DR Pfam; PF00118; Cpn60 TCP1; 1.
 DR PRINTS; PR00298; CHAPERONIN60.
 DR PRINTS; PR00304; TCOMPLEXTCP1.
 DR PROSITE; PS00750; TCP1_1; 1.
 DR PROSITE; PS00751; TCP1_2; 1.
 DR PROSITE; PS00995; TCP1_3; 1.
 KW ATP-binding; Chaperone.
 SQ SEQUENCE 563 AA; 60194 MW; 51022259612D8056 CRC64;

Query Match

Best Local Similarity 8.8%; Score 259.5; DB 2; Length 563;

Pred. No. 2.9e-11;

Matches 120; Conservative 89; Mismatches 218; Indels 83; Gaps 18;
 Qy 29 LKBIIVTSCVPSGRKLQHLNKGEGYCTTSSQALLSHLLVTHPIKILKTASTQNHVSSF 88
 Db 33 LAEMLRSSLGPKGLDKMLIDSGFD-VTTNDGATIVKEMEIQHPAAKLVEAAKAQDSEV 91
 Qy 89 SDGGLFTAILCCNL-IENVQRL---GLTPTTVIR-----LNKHL-----LSLCISYLK 132
 Db 92 GD-GTTSVAVLAGLFLKAEASLVQNIHPTIIIEGPKAFNKSLELLPOLATKVDYSDLN 150
 Qy 133 SDTCGRIPVDFSTQILLCLVRSILTSKPAKMLTRKETEHVSALILRAPLTIPENAEG 192
 Db 151 SAT-----ARDALKKIVVTTMSSK--FMAEGEELNKMINDIVIDAVTTVAEPLPDG 198
 Qy 193 --HIIIGKSLIVPLKQGVDSVTLPGILIEHSEVQMLRLLPIKKSTALKVALFCTTSLG 250
 Db 199 GYNVSLDLIKDKKGGTIEDSQLIRGIVLDKEVHAGMPRRVEKA---KIVLDASLEV 255
 Qy 251 DTSDTGEGTVVSVGVSLEN-----AVLDQ-----LNLGRQLISDHVLDVLCQKVIHPSL 301
 Db 256 EKPE-----ISAKISITSPDQIKAFIDEEAKYLDKMDVKLASIGANVVICQKGDIDIA 308
 Qy 302 KQFLNMRHIIAIDRIGVTLMPLTKMTGTQPIGSLGICPNSYGVSKVDCTAKGSOH-- 359
 Db 309 QHFLAKRGILAVRRVKRSIDIEKLEKALGARIISIKDATPEDIGYAEVVEERKVGNDKMV 368
 Qy 360 FFHLIPNEATICSILLCNENDTAWDELKLTCTQTAHVLQTLKEPWALLGGCTETHLAA 419
 Db 369 FIEGAKNPKAV-NILLRGSDMVLDEAERSINDALYSRLNLMPEYIVFGGGAIELELSA 427
 Qy 420 YIRKHTNDPESILKDECTQTELQIAEAFCSALESVVGSLHDCG-----EILTDMKYG 475
 Db 428 RLREYARS-----VGKKEQLAIEAYADALEEIPWLAETAGLAPISALMDLRAR 476
 Qy 476 HLMSVQADSPCVANWPDLLSCQCGGLYNSQ 505
 Db 477 HARG-----LTNGGVVDVINGK 492

RESULT 12

THSB_METH
 ID THSB_METH STANDARD; PRT; 538 AA.
 AC O26885;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Thermosome beta subunit (Thermosome subunit 2) (Chaperonin beta subunit).
 DE Name=thSB; OrderedLocusNames=MTH794;
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 OX NCBI_TaxID=187420;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Delta H;
 RX MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RL J. Bacteriol. 179:7135-7155(1997).
 CC -!- FUNCTION: Molecular chaperone; binds unfolded polypeptides in
 CC vitro, and has a weak ATPase activity (By similarity).
 CC -!- SUBUNIT: Forms a hetero-oligomeric complex of two stacked eight-
 CC membered rings (By similarity).
 CC -!- SIMILARITY: Belongs to the TCP-1 chaperonin family.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; AB000857; AAB85294.1; -;
 CC PIR; H69205; H69205.
 CC HSP; P48424; 1A6D.
 CC InterPro; IPR002194; Chaperonin_TCP-1.
 CC InterPro; IPR001844; Chaprinin_Cpn60.
 CC InterPro; IPR002423; Cpn60/TCP-1.
 CC InterPro; IPR008950; GroEL-ATPase.
 CC Pfam; PF00118; Cpn60 TCP1; 1.
 CC PRINTS; PR00298; CHAPERONIN60.
 CC PRINTS; PR00304; TCOMPLEXTCP1.
 CC PROSITE; PS00750; TCP1_1; 1.
 CC PROSITE; PS00751; TCP1_2; 1.
 CC PROSITE; PS00995; TCP1_3; 1.
 CC KW ATP-binding; Chaperone; Complete proteome.
 CC SEQUENCE 538 AA; 57885 MW; 493CC28C552461E3 CRC64;

Query Match 8.7%; Score 257.5; DB 1; Length 538;
 Best Local Similarity 24.1%; Pred. No. 3.9e-11;
 Matches 115; Conservative 84; Mismatches 214; Indels 65; Gaps 19;

QY 17 LTTRVTRTLVLKRIVTCYGPSGRLLKQHLNFGVGVCTTSQSSALLSHLLVTHPIIKI 76
 DB 33 ILAETVRTL-----GPKGMDKMLVDSLDIV-VTNGVILKEMDIEHPAAKM 80
 QY 77 LTASIQNHVSSFDGCLFTAILCNLI-----ENVQRIGLPPTTVIRLNKHLISLCISYLK 132
 DB 81 LVEVAKTQDEVDG-GTTTAVITAGELLKKAENLLEMEIHP-TIIAMGYRQAAEKAQEIL 138
 QY 133 SDTCGCRIPVDFSTQILLCLVRSILTSKPCMLTRKETHSVSALILRAFLLTIPENAE- 191
 DB 139 DD-----IAIDASDRDTLMKVAMTMTGK-----TEKAREPLAELIVDA-VKQVEDGEV 188
 QY 192 --GHILGKSLIVPLKQORVIDSTVLPGLIEMSEVQLMRLPIKKTALKVALFCTTLS 249
 DB 189 EKDHIEKK-----EGAAVDDSTLVQGVILDKERVH--PGMP-KKVENAKIALLNCPI- 239
 QY 250 GDTSDTGEFTVVSYGVSLENALVDQLNLGRQLISDHD-----LVLCQKVIHPSLKQFL 305
 DB 240 -EVKETEVDABEIRITDPSQMAFTEQEQEIMRDMVNSIVDTGANVLFQCKGIDDLAQHYL 298
 QY 306 NMHRIIDRIGVTLMPLTKMTGTQPIGSLGICPNYSYGVKDVCTAKE-GSQHPFHLLI 364
 DB 299 AKAGVLAVRRYKSDMEKLSKATGANIVTNIEDLSPEDLGEAGVSEKKISGSEMIFFVEE 358
 QY 365 PNEATICSLLNCRNDTAWDELKLTCTALHVLQLTLPKWPALLGGGCTETHLAAYIRHK 424
 DB 359 CKPKAVTILVRGSTEHVSEVERAIEDAIGVVAATVEDKGVAGGAPAEIETA-----412
 QY 425 THNDPESILKD--DECTQTELQLTAEAFCSALESVWGSLEHDDG-----EILTDMKYGH 476
 DB 413 -----KRLKDYADSDISGRE-QLAVSAFAEALEIVPKTLAENAGLSIDVLVLDLRAAH 463

RESULT 13
 THSA SULSH STANDARD; PRT; 560 AA.
 AC P46219;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 05-JUN-2004 (Rel. 44, Last annotation update)
 DE Thermosome alpha subunit (Thermosome subunit 1) (Chaperonin alpha subunit) (Thermophilic factor 55 alpha) (TF55-alpha) (ring complex alpha subunit) (Thermophilic factor 56).
 GN Name=chaA; Synonyms=tf56;
 OS Sulfolobus shibatae.

OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OX NCBI_TaxID=2286;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=ATCC 51178 / DSM 5389 / B12;
 RX MEDLINE=96074780; PubMed=7473746;
 RA Kagawa H.K., Osipiuk J., Maltsev N., Overbeek R., Quate-Randall E.,
 RA Joachimiak A., Trent J.D.;
 RT "The 60 kDa heat shock proteins in the hyperthermophilic archaeon
 RT Sulfolobus shibatae.";
 RL J. Mol. Biol. 253:712-725 (1995).
 CC -!- FUNCTION: Molecular chaperone; binds unfolded polypeptides in
 CC vitro, stimulates protein folding and has ATPase activity.
 CC -!- SUBUNIT: Forms a hetero-oligomeric complex of two stacked nine-
 CC membered rings; one of alpha and the other of beta subunits.
 CC -!- INDUCTION: By heat shock.
 CC -!- SIMILARITY: Belongs to the TCP-1 chaperonin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; L34691; AAB87624.1; -;
 CC PIR; S59859; S59859.
 CC HSP; P48424; 1A5S.
 CC InterPro; IPR002194; Chaperonin_TCP-1.
 CC InterPro; IPR001844; Chaprinin_Cpn60.
 CC InterPro; IPR002423; Cpn60/TCP-1.
 CC InterPro; IPR008950; GroEL-ATPase.
 CC Pfam; PF00118; Cpn60 TCP1; 1.
 CC PRINTS; PR00298; CHAPERONIN60.
 CC PRINTS; PR00304; TCOMPLEXTCP1.
 CC PROSITE; PS00750; TCP1_1; 1.
 CC PROSITE; PS00751; TCP1_2; 1.
 CC PROSITE; PS00995; TCP1_3; 1.
 CC KW ATP-binding; Chaperone; Direct protein sequencing; Heat shock.
 CC SEQUENCE 560 AA; 59703 MW; F42C945E74565604 CRC64;

 Query Match 8.7%; Score 256.5; DB 1; Length 560;
 Best Local Similarity 24.3%; Pred. No. 4.9e-11;
 Matches 115; Conservative 84; Mismatches 221; Indels 53; Gaps 16;
 QY 29 LKRVITSCYGPSGRLLKQHLNFGVGVCTTSQSSALLSHLLVTHPIIKILTASIQNHVSSP 88
 DB 31 LAEMLRSSLGPKGLDKMLDSFGD-VTIINDGATIVKDMEIQHPPAKLLIVEAKAQDAEV 89
 QY 89 SDGGLFTAILCCNLIENVORL---GLTPTTVIRLNKHLISLCISYLVKSDTCGCRIPVDFS 145
 DB 90 GDGTTSAVVLAGALLEKAESLLDQNIHTPTIIIEGYKAVTKALELLPQ--LGTRIDIRD 147
 QY 146 STQILLCLVRSI-----LTSKPACMLTRKETHSVSALILRAFLLTIP---NAEGHILGK 198
 DB 148 NSSVARDTLRKIAFTTTLASK--FIAEGABELNKIIDMVIDA-IVNVAEPLPNGSYNVSLDL 204
 QY 199 SLIVPLKQORVIDSTVLPGLIEMSEVQLMRLPIKKTALKVALFCTTLSGTSSTGEG 258
 DB 205 IKIDKKKGSIEDSVLVKGLVDKEVVH--PGMP-PRVTAKAVLDALEVEKPE----257
 QY 259 TVVVSYGVSLEN-----AVLDQ---LLNLGRQLISDHDVLVLCQKVIHPSLKQFLNMHR 309
 DB 258 ---ISAKISITSPQIKAFIDBESKYLKDMVDKIASIGANVVICQKGIIDDLAQHFLAKKG 314
 QY 310 ITAIDRIGVTLMPLTKMTGTQPIGSLGICPNYSYGVKDVCTAKEGSOH--FFHLIPNE 367
 DB 315 ILAVRRYKSDMEKLEKALGARIISIKDPTDGLGYAEIIEVRRVGNOKMVFIEGAKNL 374
 QY 368 ATICSELLNCRNDTAWDELKLTCTALHVLQLTLPKWPALLGGGCTETHLAAYIRKTHN 427

Db 375 KAV-NILGRSNDMALDEAERSINDALHARNILLEVLPVILPGGGAIELEAMKLREYARS 433
QY 428 DPESILKDDCTQTELOLIAEAFCSALESVVGSLHDGG-----EILTDMKYGH 476
Db 434 -----VGKKEQLAEAFADALEEPTILAEATAGLEAISALMDLRARH 475

RESULT 14

THSA SULAC
ID THSA SULAC STANDARD; PRT; 490 AA.
AC Q9V2T5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE Thermosome alpha subunit (Thermosome subunit 1) (Chaperonin alpha subunit) (Thermophilic factor 55 alpha) (TF55-alpha) (Fragment)
GN Name:thsa;
OS Sulfolobus acidocaldarius.
OC Archaea; Euryarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2285;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=9439900; PubMed=10508614; DOI=10.1016/S0960-9822(99)80457-6;
RA Archibald J.M., Logsdon J.M. Jr., Doolittle W.F.;
RT "Recurrent paralogy in the evolution of archaeal chaperonins.";
RL Curr. Biol. 9:1053-1056(1999).
CC -1- FUNCTION: Molecular chaperone; binds unfolded polypeptides in vitro, and has a weak ATPase activity (By similarity).
CC -1- SUBUNIT: Forms a hetero-oligomeric complex of two stacked eight-membered rings (By similarity).
CC -1- SIMILARITY: Belongs to the TCP-1 chaperonin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF149923; AAF03364.1; -.
DR HSSP; P48424; IASS.
DR InterPro; IPR002194; Chaperonin TCP-1.
DR InterPro; IPR001844; Chaprin Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; Cpn60 TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00750; TCP1_1; 1.
DR PROSITE; PS00751; TCP1_2; 1.
DR PROSITE; PS00995; TCP1_3; 1.
KW ATP-binding; Chaperone.
FT NON TER 1 1
FT NON TER 490 490
SQ SEQUENCE 490 AA; 52843 MW; 07C0A186600AE868 CRC64;

Query Match

Best Local Similarity 8.6%; Score 255; DB 1; Length 490;
Matches 107; Conservative 90; Mismatches 219; Indels 50; Gaps 14;

QY 35 SCYGPSRLKQLHNGFGYVCTTSQSALLSHLVTHPIKILTAISONHVVSFSCGLF 94
Db 3 SSIGPRLDKMLDSDFGD-VTIYNDGATIVKMEIOHPAAKLIVAEAKQAQDAEVDG-GTT 60
QY 95 TAILCCNLI-----ENVORGLTPTTVIRLNKHLISLCISYKSDTCGCRIPVDFSSQTIL 150
Db 61 SAVVLACGLLDKAEELLEQNVHPHPIIDGYKKALTKALEII--DQLSLKIDVNDLSPTA 118
QY 151 LCLVRSILTSKPAKMLTRKTEHVSALI---LRAFLTTPENAE--HIILGKSLIVPK 205
Db 119 KAQLKKIVTSMSSKLIAGAEIDKIIDLIAAITVAEKRPDGTYNVPLDLIKIDKKK 178

QY 206 GORVIDSTVLPGLIEMSEVOLMRLLPIKKTALKKVALFCTTSLGSDTSDTGEVTVVSYG 265
Db 179 GGSIEISILVHGLVLDKEVHVHAGMPRVEKA---KIAVLDAALEVEKPE-----ISAK 228
QY 266 VSIEN-----AVLDQ-----LLNLGRQLISDHVLDVLCQKVIHPSLKFOLNMHRIIADRI 316
Db 229 ISITSPEQIKSFLEDEARVLEKEMVDKLASIGANVVICQKIGIDVVAQHLAKKGLAVRRV 288
QY 317 GVTLMPELTKMTCTQPIGSLGSIKPNISYGVKDVCTAKGSOH--PFHLIPNEATICSLL 374
Db 289 KRSDIEKLEKALGARIISIKDATPEDLGYAEVLVEERRIGNDKWVFEGAKNPRV-NIL 347
QY 375 LCNRNNTAWDELKLTCTQALHVLQTLKEPWALLGGCTETHLAAYTRHKTHNDPESILK 434
Db 348 LRGSNDMALDEAERSLNDALHSRLNVLMKPMIVAGGAVSESELALRLREYARS----- 400
QY 435 DDECTQTEIQLIABAFCSALESVVGSLHDGG-----EILTDMKYGH 476
Db 401 ----VGKKEQLAEKFAALEEIPMILAEATAGMEPIQALMDLRARH 442

RESULT 15

Q6LI32
ID Q6LI32 PRELIMINARY; PRT; 546 AA.
AC Q6LI32;
DT 05-JUL-2004 (TEMBLrel. 27, Created)
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TEMBLrel. 27, Last annotation update)
DE Thermosome subunit.
DE OrderedLocusNames=PTO0735;
OS Picrophilus torridus.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatiales;
OC Picrophilaceae; Picrophilus.
OX NCBI_TaxID=82076;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=DSM 9790 / ATCC 700027;
RX PubMed=15184674; DOI=10.1073/pnas.0401356101;
RA Fuetterer O., Angelov A., Liesegang H., Gottschalk G., Schleper C., Scheper B., Dock C., Antranikian G., Liebl W.;
RT "Genome sequence of Picrophilus torridus and its implications for life around pH 0.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:9091-9096(2004).
DR EMBL; AE017261; AAT43320.1; -.
DR HSSP; P48424; IASS.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0051082; F:unfolded protein binding; IEA.
DR GO; GO:0006457; P:protein folding; IEA.
DR InterPro; IPR002194; Chaperonin TCP-1.
DR InterPro; IPR001844; Chaprin Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; Cpn60 TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00750; TCP1_1; 1.
DR PROSITE; PS00751; TCP1_2; 1.
DR PROSITE; PS00995; TCP1_3; 1.
KW ATP-binding; Chaperone; Complete proteome.
SQ SEQUENCE 546 AA; 58820 MW; 864ED7D4CAF736A0 CRC64;

Query Match 8.6%; Score 253.5; DB 2; Length 546;
Best Local Similarity 22.6%; Pred. No. 8.1e-11;
Matches 129; Conservative 95; Mismatches 243; Indels 103; Gaps 21;

QY 20 ERVITTLVSKRIVTSCYGPSRLKQLHNGFGYVCTTSQSALLSHLVTHPIKILTA 79
Db 36 DAVRITL-----GPKGMDKMLVDSIGDIV-ITNDGATILKEMIDHPTAKLVE 83
QY 80 STONHVVSFSDCGLFTAILCCNLIENVQL---GLTPTTVIRLNKHLISLCISYKSDTC 136
Db 84 ASKSQDTAVGDGTTTWWVLAGELLKQAESLLEQGV-HSTVIASGYH---LAVTEAKQLD 139

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2005, 04:43:28 ; Search time 170 Seconds
(without alignments)
1296.785 Million cell updates/sec

Title: US-10-616-263-30

Perfect score: 2957

Sequence: 1 MSRLKAKPSLCKSEPLTTE.....VAVETANLIWLSYVIEDKN 570

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760391 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2950	99.8	570	3 AAY94912	Aay94912 Human sec
2	2936	99.3	570	3 AAY91944	Aay91944 Human cha
3	2936	99.3	572	5 ABP41775	Abp41775 Human ova
4	2904	98.2	570	7 ADJ69265	Adj69265 Human hea
5	2503	84.6	498	8 ADP43680	Adp43680 Human PMM
6	2325	78.6	474	4 ABG19636	Abg19636 Novel hum
7	2086	70.5	406	3 AABF3205	Aabf3205 Human ORF
8	1092	36.9	222	3 AAB56236	Aab56236 Human sec
9	1076	36.4	223	3 AAB56218	Aab56218 Human sec
10	841	28.4	166	3 AAB56356	Aab56356 Human sec
11	432	14.6	86	3 AAG02893	Ag02893 Human sec
12	257.5	8.7	538	8 ADN18515	Adn18515 Bacterial
13	256.5	8.7	560	7 AAO24120	Aao24120 S shibata
14	248.5	8.4	554	7 AAO24111	Aao24111 A pernix
15	248.5	8.4	557	8 ADS44430	Ads44430 Bacterial
16	245	8.3	520	7 ADB70179	Adb70179 C. neofor
17	238.5	8.1	558	6 ABG74939	Abg74939 A. Gossyp
18	238	8.0	545	8 ADS42984	Ads42984 Bacterial
19	236	8.0	530	8 ADS22203	Ads22203 Bacterial
20	235	7.9	537	8 ADN20780	Adn20780 Bacterial
21	235	7.9	542	7 AAO24112	Aao24112 M mazel a
22	232	7.8	535	7 AAO24121	Aao24121 S shibata
23	229.5	7.8	542	8 ADN18378	Adn18378 Bacterial
24	227	7.7	559	4 AAM79916	Aam79916 Human pro
25	227	7.7	567	6 ABJ25946	Abj25946 Aspergill

26	226	7.6	555	8 ADS44512	Ads44512 Bacterial
27	225.5	7.6	545	6 ABR63888	AbR63888 T acdioph
28	225.5	7.6	549	8 ADS29192	Ads29192 Bacterial
29	225	7.6	544	7 ADD47398	Add47398 Human Pro
30	225	7.6	544	8 ADQ30558	Adq30558 Pancreas
31	225	7.6	544	8 ABM80158	Abm80158 Tumour-as
32	225	7.6	545	4 AAM78932	Aam78932 Human pro
33	225	7.6	559	4 AAE13836	Aae13836 Human lun
34	225	7.6	559	7 ADD66704	Add66704 Human lun
35	225	7.6	559	7 ADE87958	Ade87958 Human lun
36	222	7.5	544	8 ADP54145	Adp54145 Human PRO
37	222	7.5	552	8 ADN18464	Adn18464 Bacterial
38	221.5	7.5	526	8 ADO57323	Ado57323 Kidney de
39	220.5	7.5	534	8 ADS21954	Ads21954 Bacterial
40	220.5	7.5	541	8 ADS24471	Ads24471 Bacterial
41	219.5	7.4	541	8 ADS21926	Ads21926 Bacterial
42	219	7.4	544	4 AAB80955	Aab80955 Viral pro
43	218	7.4	540	5 ABP73776	Abp73776 Candida a
44	218	7.4	542	8 ADS22242	Ads22242 Bacterial
45	217.5	7.4	543	7 AAO24107	Aao24107 T acidoph

ALIGNMENTS

RESULT 1

AAV94912

ID AAY94912 standard; protein; 570 AA.

AC AAY94912;

XX

XX

DT 16-JUN-2000 (first entry)

XX

DE Human secreted protein clone pp314_19 protein sequence SEQ ID NO:30.

XX

KW Human; secreted protein; immunestimulant; immunosuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antichyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy.

XX Homo sapiens.

OS

XX WO200009552-A1.

PD 24-FEB-2000.

 XX | PF | 13-AUG-1999; | 99WO-US018298. || PR | 14-AUG-1998; | 98US-0096622P. |
PR	17-AUG-1998;	98US-0096815P.
PR	04-SEP-1998;	98US-0099229P.
PR	23-OCT-1998;	98US-0105368P.
PR	08-JAN-1999;	99US-0115234P.
PR	12-FEB-1999;	99US-0119931P.
PR	18-FEB-1999;	99US-0120575P.
PR	30-APR-1999;	99US-0132020P.
PR	11-AUG-1999;	99US-0148424P.
XX	(GEMY) GENETICS INST INC.	
PA	Jacobs K, Mccoy JM, Lavallie ER, Collins-Racie LA, Evans C;	
PI	Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;	
PI	Wong GG, Clark HF, Fechtel K;	
DR	WPI; 2000-205979/18.	
PT	New polynucleotides encoding secreted proteins, which may have e.g.	

PT nutritional, chemokine, immune stimulating or suppressing, hematopoiesis
PT regulating, tissue growth, activin/inhibin antiinflammatory or tumor
PT inhibition activity.

XX
XX Claim 39; Page 499-501; 641pp; English.

CC AAA16618 to AAA16697 encode the human secreted proteins given in AAY94898
CC to AAY94980, isolated from human adult brain, adult thyroid, adult
CC retina, foetal carcinoma, adult blood, adult neural, foetal kidney, adult
CC placenta, adult testis, whole embryo, adult cartilage, kidney, foetal
CC brain, adult thymus, foetal placenta, adult uterus, adult tumour, and
CC adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans and
CC animals. The polynucleotides can be used as markers for tissues in which
CC the protein is preferentially expressed, as molecular weight markers on
CC Southern gels, and as chromosome markers or tags to identify chromosomes
CC or to map gene positions. The proteins can be used in the treatment of
CC immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention

XX Sequence 570 AA;

Query Match 99.8%; Score 2950; DB 3; Length 570;
Best Local Similarity 99.8%; Pred. No. 2.7e-297;
Matches 569; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSRLAKPKSLCKSEPLTTERVVTLSVLKRIYTCYSGPSGRUKQIHNGFGGVCTTSOS 60
DB 1 MSRLAKPKSLCKSEPLTTERVVTLSVLKRIYTCYSGPSGRUKQIHNGFGGVCTTSOS 60
QY 61 SALLSHLLVTHPKILKILTASIQNHVSFSDCGLFTAILCCNLJENVQRLGLTPTTVIRLN 120
DB 61 SALLSHLLVTHPKILKILTASIQNHVSFSDCGLFTAILCCNLJENVQRLGLTPTTVIRLN 120
QY 121 KHLISLCISYLSKSDTCGRIPVDFSSQIILLCLVRSILTSKPACMLTRKETEHSALILR 180
DB 121 KHLISLCISYLSKSDTCGRIPVDFSSQIILLCLVRSILTSKPACMLTRKETEHSALILR 180
QY 181 AFLTTPENAEGLIILGKSLIIVPLKQORVIDSTVLPGLIEMSEVQIMRLPIPKKTALK 240
DB 181 AFLTTPENAEGLIILGKSLIIVPLKQORVIDSTVLPGLIEMSEVQIMRLPIPKKTALK 240
QY 241 VALFCTTSLGSDTSDGTGVTWVSYGSLENAVLDQLNLGRQLISDHVDLVLCQKVIHPS 300
DB 241 VALFCTTSLGSDTSDGTGVTWVSYGSLENAVLDQLNLGRQLISDHVDLVLCQKVIHPS 300
QY 301 LKQFLNMHRIIADRIQVTLMELPTKMTGTPQPIGSLGICPNYSYGVKDVCTAKFGSQHF 360
DB 301 LKQFLNMHRIIADRIQVTLMELPTKMTGTPQPIGSLGICPNYSYGVKDVCTAKFGSQHF 360
QY 361 FHLIPNEATCSLLLCNRNDTANDELKLTCTQTAHVLTQLKEPWLIGGGCTETHLAAY 420
DB 361 FHLIPNEATCSLLLCNRNDTANDELKLTCTQTAHVLTQLKEPWLIGGGCTETHLAAY 420
QY 421 IRKHTNDPESILKDBDCTCTELQLIAEAFCSALESVVGSLHDGGGELLTDMKYGHLWSV 480
DB 421 IRKHTNDPESILKDBDCTCTELQLIAEAFCSALESVVGSLHDGGGELLTDMKYGHLWSV 480
QY 481 QADSPCVANWPDLLSQCGCGLYNSQBELNWSFLRSTRPPVPOCLPHEAVGASNLITLD 540
DB 481 QADSPCVANWPDLLSQCGCGLYNSQBELNWSFLRSTRPPVPOCLPHEAVGASNLITLD 540

QY 541 CLTAKLSGLQAVAVETANLIWLSYVIEDKN 570
DB 541 CLTAKLSGLQAVAVETANLIWLSYVIEDKN 570

RESULT 2
AAY91944
ID AAY91944 standard; protein; 570 AA.
XX
AC AAY91944;
XX 19-JUL-2000 (first entry)
XX Human chaperone protein 5 (HCHP-5).
DE Human chaperone protein; HCHP-5; neurodegenerative; cancer; metabolic;
KW developmental; autoimmune; inflammatory; cell proliferation.
XX Homo sapiens.
XX Key Location/Qualifiers
FH Modified-site 2
FT /note= "potential phosphorylation site"
FT Modified-site 10
FT /note= "potential phosphorylation site"
FT Modified-site 19
FT /note= "potential phosphorylation site"
FT Modified-site 40
FT /note= "potential phosphorylation site"
FT Modified-site 87
FT /note= "potential phosphorylation site"
FT Modified-site 159
FT /note= "potential phosphorylation site"
FT Modified-site 167
FT /note= "potential phosphorylation site"
FT Modified-site 171
FT /note= "potential phosphorylation site"
FT Modified-site 185
FT /note= "potential phosphorylation site"
FT Binding-site 192..199
FT /label= ATP/GTP-binding_site_motif_A
FT /note= "P loop"
FT Modified-site 237
FT /note= "potential phosphorylation site"
FT Modified-site 300
FT /note= "potential phosphorylation site"
FT Modified-site 319
FT /note= "potential phosphorylation site"
FT Modified-site 346
FT /note= "potential phosphorylation site"
FT Modified-site 352
FT /note= "potential phosphorylation site"
FT Modified-site 379
FT /note= "potential glycosylation site"
FT Modified-site 381
FT /note= "potential phosphorylation site"
FT Modified-site 400
FT /note= "potential phosphorylation site"
FT Modified-site 425
FT /note= "potential phosphorylation site"
FT Modified-site 439
FT /note= "potential phosphorylation site"
FT Modified-site 452
FT /note= "potential phosphorylation site"
FT Modified-site 504
FT /note= "potential glycosylation site"
FT Modified-site 509
FT /note= "potential glycosylation site"
FT Modified-site 515
FT /note= "potential phosphorylation site"
FT Modified-site 516
FT /note= "potential phosphorylation site"
FT Modified-site 536

PT	Modified-site	/note= "potential glycosylation site"	
FT	543		
FT	/note= "potential phosphorylation site"		
XX	W0200017358-A2.		
XX	30-MAR-2000.		
XX	22-SEP-1999;	99WO-US022027.	
XX	22-SEP-1998;	98US-00159642.	
PR	22-SEP-1998;	98US-0172221P.	
PR	19-JAN-1999;	99US-00233291.	
PR	19-APR-1999;	99US-00294698.	
PR	19-APR-1999;	99US-0172232P.	
XX	(INCY-) INCYTE PHARM INC.		
PA	Tang YT, Hillman JL, Yue H, Patterson C, Baughn MR, Batra S;		
XX	WPI; 2000-283583/24.		
DR	N-PSDB; AAA08567.		
XX	New purified polypeptides and polynucleotides encoding human chaperone		
PT	proteins, useful for diagnosing, treating and preventing disorders		
PT	associated with the expression human chaperone proteins.		
XX	Claim 1; Page 73-75; 88pp; English.		
PS	AA91940-15 are human chaperone proteins 1-6 (HCHP-1 to HCHP-6)		
CC	respectively. The sequences can be used to treat and prevent disorders		
CC	associated with altered expression or activity of HCHP comprising		
CC	administering a composition comprising the polypeptide or an antagonist		
CC	to a patient (claimed). The human chaperone proteins are also useful for		
CC	the diagnosis, treatment or prevention of neurodegenerative, metabolic,		
CC	developmental, autoimmune/inflammatory disorders and cell proliferative		
CC	disorders including cancer		
XX	Sequence 570 AA;		
SQ	Query Match	99.3%; Score 2936; DB 3; Length 570;	
	Best Local Similarity	99.5%; Pred. No. 7.8e-296;	
	Matches 567; Conservative	2; Mismatches 1; Indels 0; Gaps 0;	
QY	1 MSRLKAKPKLCKSEPLTTERVTTLSVLKRIVTSCVPSGRLLKQLHNGFGVVCITTSOS	60	
Db	1 MSRLKAKPKLCKSEPLTTERVTTLSVLKRIVTSCVPSGRLLKQLHNGFGVVCITTSOS	60	
QY	61 SALLSHLLVTHPIKILITASIQNHVSSFSDCGLFTAILCCNLJENQRLGLTPTTVIRLN	120	
Db	61 SALLSHLLVTHPIKILITASIQNHVSSFSDCGLFTAILCCNLJENQRLGLTPTTVIRLN	120	
QY	121 KHLISICISYKSDTCGRIPVDFSSQIILLVRSILTSKPCMLTRKETEHSALILR	180	
Db	121 KHLISICISYKSDTCGRIPVDFSSQIILLVRSILTSKPCMLTRKETEHSALILR	180	
QY	181 AFLTTPPENAEGLIILGKSLIVPLKQGVDSITVPLGILTIEMSEVOLMRLLPKKSTALK	240	
Db	181 AFLTTPPENAEGLIILGKSLIVPLKQGVDSITVPLGILTIEMSEVOLMRLLPKKSTALK	240	
QY	241 VALFCTTSLGSDTSGEGTVVSVGYSLNVLNQLNLGRQLISDHVDLVLCKQVIHPS	300	
Db	241 VALFCTTSLGSDTSGEGTVVSVGYSLNVLNQLNLGRQLISDHVDLVLCKQVIHPS	300	
QY	301 LKQFLNMRHIIADRGVTLMELPTKMTGTQPGISGICPNISYGVSKDVCTAKFGSKHF	360	
Db	301 LKQFLNMRHIIADRGVTLMELPTKMTGTQPGISGICPNISYGVSKDVCTAKFGSKHF	360	
QY	361 FHLIPNEATICSLLCNRNDDTANDELKLTCTQTAHVLTQLTKEPWALLGGGCTETHLAY	420	
Db	361 FHLIPNEATICSLLCNRNDDTANDELKLTCTQTAHVLTQLTKEPWALLGGGCTETHLAY	420	
QY	421 IRKHTNDPESILKDDCTQTTELQIAEAFCSALESVVGSLHEDGGEILLTDMKYGHLWSV	480	

Db	421	IRKHTNDPESILKDDCTQTTELQIAEAFCSALESVVGSLHEDGGEILLTDMKYGHLWSV	480
QY	481	QADSPCVANWPDLLSQCGCLYNSQBELNWSFLSTRRPFVPSQCLPHEAVGSASNLTLTD	540
Db	481	QADSPCVANWPDLLSQCGCLYNSQBELNWSFLSTRRPFVPSQCLPHEAVGSASNLTLTD	540
QY	541	CLTAKLSGLQVAVETANLIWDLISYVIEDKN	570
Db	541	CLTAKLSGLQVAVETANLIWDLISYVIEDKN	570
RESULT 3			
ABP411775			
ID	ABP41775	standard; protein; 572 AA.	
XX	AC	ABP41775;	
XX	XX		
DT	22-AUG-2002	(first entry)	
XX	XX	Human ovarian antigen HMAEL73, SEQ ID NO:2907.	
DE	XX		
XX	XX	Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;	
KW	ovarian cancer; breast cancer; tumour; reproductive system disorder;		
KW	infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;		
KW	PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;		
KW	inflammatory condition; immune disorder; blood disorder;		
KW	cardiovascular disorder; respiratory disorder; neurological disorder;		
KW	gastrointestinal disorder; urinary system disorder; drug screening;		
KW	gene therapy; chromosome mapping; forensic analysis;		
KW	antibody preparation; cytostatic; immunomodulatory; neuroprotective;		
KW	antiinflammatory; gynaecological; reproductive.		
XX	XX	Homo sapiens.	
OS	XX		
XX	XX	WO200200677-A1.	
PN	XX		
XX	XX	03-JAN-2002.	
XX	XX		
PF	07-JUN-2001;	2001WO-US018569.	
XX	XX		
PR	07-JUN-2000;	2000US-0209467B.	
XX	XX	(HUMA-) HUMAN GENOME SCI INC.	
PA	XX		
XX	XX	Biree CE, Rosen CA;	
PI	XX		
XX	XX	WPI; 2002-147878/19.	
DR	XX	N-PSDB; ABQ54852.	
PT	Isolated nucleic acid molecules encoding novel ovarian polypeptides,		
PT	useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian		
PT	cancer), immune disorders, cardiovascular disorders and neurological		
PT	diseases.		
XX	XX	Claim 11; SEQ ID NO 2907; 2922pp; English.	
PS	XX		
XX	XX	The invention relates to 2175 novel human ovarian antigens (ABP41054-	
CC	ABP43228) and to cDNAs encoding them (ABQ5431-ABQ56305), and also		
CC	encompasses polypeptides 90% identical and polynucleotides 95% identical		
CC	to the sequences of the invention. The invention additionally relates to		
CC	recombinant vectors and host cells comprising human ovarian antigen		
CC	polynucleotides, antibodies against human ovarian antigens, and the use		
CC	of ovarian antigen polynucleotides and polypeptides in diagnosing,		
CC	treating, prognosing or preventing various ovary and/or breast-related		
CC	disorders. Such conditions include ovarian cancer and breast cancer, and		
CC	metastatic tumours of ovarian or breast origin, reproductive system		
CC	disorders (e.g., infertility, disorders of pregnancy, anovulation,		
CC	polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine		
CC	disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic		
CC	shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and		
CC	vaginitis), immune disorders (e.g., congenital and acquired		
CC	immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),		

CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences
XX
SQ Sequence 572 AA;

Query Match 99.3%; Score 2936; DB 5; Length 572;
Best Local Similarity 99.5%; Pred. No. 7.8e-296;
Matches 567; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MSRLKAKPSLCKSEPLTTERVTTLSVLKRVITSCYGPSGRKLQHLNFGGVCTTSQS 60
DB 3 MSRLKAKPSLCKSEPLTTERVTTLSVLKRVITSCYGPSGRKLQHLNFGGVCTTSQS 62
QY 61 SALLSHLLVTHPILKILITASIQNHVSSFSDCGLFTAILCCNLNIENVQRLGTPPTVIRLN 120
DB 63 SALLSHLLVTHPILKILITASIQNHVSSFSDCGLFTAILCCNLNIENVQRLGTPPTVIRLN 122
QY 121 KHLLSLCISYLKSDTCGCRIPVDFSTQILLCLVRSILTSKPACMLTRKETHVSALIIR 180
DB 123 KHLLSLCISYLKSDTCGCRIPVDFSTQILLCLVRSILTSKPACMLTRKETHVSALIIR 182
QY 181 AFLTTPENAEAGHIIIGKSLIVPLKQORVIDSTVLPGLIEMSEVOLMRLPIKKTALK 240
DB 183 AFLTTPENAEAGHIIIGKSLIVPLKQORVIDSTVLPGLIEMSEVOLMRLPIKKTALK 242
QY 241 VALFCTTSLGSDTSGEGTVVSYGVSLENVAVDQLNLGRQLISDHVDLVLCQKVIHPS 300
DB 243 VALFCTTSLGSDTSGEGTVVSYGVSLENVAVDQLNLGRQLISDHVDLVLCQKVIHPS 302
QY 301 LKQFLNMRHIIADRIQVTLMEPLTKMTGTQPIGSLGSIICPNISYGVSKDVCTAKFGSQHF 360
DB 303 LKQFLNMRHIIADRIQVTLMEPLTKMTGTQPIGSLGSIICPNISYGVSKDVCTAKFGSKHF 362
QY 361 FHLIPNEATCSLLCNRNNTAWDELKLTQCTALHVLQTLKPSWALLGGCGCTETHLAAY 420
DB 363 FHLIPNEATCSLLCNRNNTAWDELKLTQCTALHVLQTLKPSWALLGGCGCTETHLAAY 422
QY 421 IRKHTNDPESILKDDCTCTELQIAEAFCSALESVVGSLHDGGEILTDMKYGHWSV 480
DB 423 IRKHTNDPESILKDDCTCTELQIAEAFCSALESVVGSLHDGGEILTDMKYGHWSV 482
QY 481 QADSPCVANWPDLLSQCGCLYNSQBELNWSFLRSTRRPVPOSCLPHEAVGASNLITLD 540
DB 483 QADSPCVANWPDLLSQCGCLYNSQBELNWSFLRSTRRPVPOSCLPHEAVGASNLITLD 542
QY 541 CLTAKLSGLQVAVETANLIWDSYVIEDKN 570
DB 543 CLTAKLSGLQVAVETANLIWDSYVIEDKN 572

RESULT 4
ADJ69265
ID ADJ69265 standard; protein; 570 AA.
XX
AC ADJ69265;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human heat mitochondrial protein as a therapeutic target SeqID1071.
XX
KW mitochondrial; human; screening assay; diabetes mellitus;
KW Huntington's disease; osteoarthritis;

KW Leber's hereditary optic neuropathy; LHON;
KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
KW neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;
KW osteopathic; ophthalmological; cytostatic.
OS Homo sapiens.
XX
XX WO2003087769-A2.
XX 23-OCT-2003.
XX
XX 04-APR-2003; 2003WO-US010870.
XX
XX 12-APR-2002; 2002US-0372843P.
XX 17-JUN-2002; 2002US-0389987P.
XX 20-SEP-2002; 2002US-0412418P.
XX
XX (MITO-) MITOKOR.
XX (BUCK-) BUCK INST AGE RES.
XX
XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
XX Warnock DE;
XX
XX WPI; 2003-845369/78.
DR
XX
XX
PT Identifying a mitochondrial target for drug screening assays and for
PT treating diseases associated with altered mitochondrial function,
PT comprises detecting a modified polypeptide in a sample and correlating
PT with the disease.
XX
XX Claim 1; SEQ ID NO 1071; 180pp; English.
PS
XX
CC This invention relates to novel mitochondrial targets that can be used
CC for therapeutic intervention in treating a disease associated with
CC altered mitochondrial function. Specifically, it refers to a method for
CC identifying proteins of the human heart mitochondrial proteome that are
CC useful for drug screening assays, as well as therapeutic targets. The
CC present invention describes a method for identifying such proteins that
CC can be used in the treatment of various diseases associated with altered
CC mitochondrial function including diabetes mellitus, Huntington's disease,
CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
CC compositions have neuroprotective, nootropic, antidiabetic,
CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
CC cytostatic activities. This polypeptide sequence is a human heart
CC mitochondrial protein of the invention.
XX
SQ Sequence 570 AA;

Query Match 98.2%; Score 2904; DB 7; Length 570;
Best Local Similarity 98.6%; Pred. No. 1.7e-292;
Matches 562; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
QY 1 MSRLKAKPSLCKSEPLTTERVTTLSVLKRVITSCYGPSGRKLQHLNFGGVCTTSQS 60
DB 1 MSRLKAKPSLCKSEPLTTERVTTLSVLKRVITSCYGPSGRKLQHLNFGGVCTTSQS 60
QY 61 SALLSHLLVTHPILKILITASIQNHVSSFSDCGLFTAILCCNLNIENVQRLGTPPTVIRLN 120
DB 63 SALLSHLLVTHPILKILITASIQNHVSSFSDCGLFTAILCCNLNIENVQRLGTPPTVIRLN 120
QY 121 KHLLSLCISYLKSDTCGCRIPVDFSTQILLCLVRSILTSKPACMLTRKETHVSALIIR 180
DB 123 KHLLSLCISYLKSDTCGCRIPVDFSTQILLCLVRSILTSKPACMLTRKETHVSALIIR 180
QY 181 AFLTTPENAEAGHIIIGKSLIVPLKQORVIDSTVLPGLIEMSEVOLMRLPIKKTALK 240
DB 183 AFLTTPENAEAGHIIIGKSLIVPLKQORVIDSTVLPGLIEMSEVOLMRLPIKKTALK 240
QY 241 VALFCTTSLGSDTSGEGTVVSYGVSLENVAVDQLNLGRQLISDHVDLVLCQKVIHPS 300

Db 241 VALFCTTLTGSDTSGTGVVSVGVLENVAVDQLNLGRQLISDHVDLVLQKVIHPS 300
Qy 301 LKQFLNMHRIIAIDRIGVTLMPELTKMTGTQPIGSLGSI CPNSVSGVKDVTAKFGSOHF 360
Db 301 LKQFLNMHRIIAIDRIGVTLMPELTKMTGTQPIGSLGSI CPNSVSGVKDVTAKFGSKHF 360
Qy 361 FLHIPNEATCSLLLCNRNDTAMDELKLTCTQALHVLQTLKBPWALLGGGCTETHLAAY 420
Db 361 FLHIPNEATCSLLLCNRNDTAMDELKLTCTQALHVLQTLKBPWALLGGGCTETHLAAY 420
Qy 421 IRKTHNDPSSILKDDCTQTELQIAEAFCSALESVVGSI EHDGGEILTDKMYGHLWSV 480
Db 421 IRKTHNDPSSILKDDCTQTELQIAEAFCSALESVVGSI EHDGGEILTDKMYGHLWSV 480
Qy 481 QADSPCVANWPDLLSQCGGLYNSQBELNWSFLRSTRPVPQSCLPHEAVGASNLTLTD 540
Db 481 HADSPCVANWPDLLSQCGGLYNSQBELNWSFLRSTRPVPQSCLPHEAVGASNLTLTD 540
Qy 541 CLTAKLSGLQVAVETANLIWDLVSIVIEDKN 570
Db 541 CLTAKLSGLQVAVETANLIWDLVSIVIEDKN 570

RESULT 5
ADP43680
ID ADP43680 standard; protein; 498 AA.
XX
AC ADP43680;
XX
DT 18-NOV-2004 (first entry)
DE Human PMMM-37 protein SEQ ID NO:37.
XX
KW human; protein modification and maintenance molecule; PMMM;
KW gastrointestinal; cardiovascular; immunosuppressive; antiinflammatory;
KW cytoskeletal; neuroprotective; gynaecological; gene therapy;
KW gastrointestinal disorder; cardiovascular disorder; autoimmune disorder;
KW inflammatory disorder; cell proliferative disorder;
KW developmental disorder; epithelial disorder; neurological disorder;
KW reproductive disorder.
XX
OS Homo sapiens.
XX
XX WO2004053068-A2.
XX
PD 24-JUN-2004.
XX
PF 03-DEC-2003; 2003WO-US038573.
XX
PR 05-DEC-2002; 2002US-0431639P.
PR 17-DEC-2002; 2002US-0434315P.
PR 24-JAN-2003; 2003US-0442442P.
PR 30-JAN-2003; 2003US-0444141P.
PR 21-FEB-2003; 2003US-0449491P.
XX
PA (INCY-) INCYTE CORP.
XX
XX Becha SD, Hafalia AJA, Swarnakar A, Rankumar J, Richardson TW,
PI Kable AE, Marquis JP, Khare R, Lee SY, Tran UK, Bhatia UG;
PI Burrill JD, Blake JJ, Ho A, Zheng W, Gao J, Chawla NK, Mason PM;
PI Jin P, Lee S;
XX
DR WPI; 2004-468843/44.
DR N-PSDB; ADP43721.
XX
XX New human protein modification and maintenance molecules and
PT polynucleotides for diagnosing, preventing or treating diseases
PT associated with aberrant protein expression, e.g. cardiovascular or cell
PT proliferative disorders.
XX
PS Claim 1; SEQ ID NO 37; 268pp; English.
XX
XX The present sequence represents a human protein modification and

CC maintenance molecule (PMMM) (1), designated PMMM-37. Also described: (1)
CC an isolated polynucleotide encoding (1); (2) a recombinant polynucleotide
CC comprising a promoter sequence operably linked to the polynucleotide in
CC (1); (3) a cell transformed with the recombinant polynucleotide; (4) a
CC transgenic organism comprising the recombinant polynucleotide; (5)
CC methods of producing or purifying (1); (6) an isolated antibody that
CC specifically binds to (1); (7) detecting a target polynucleotide or (1)
CC in a sample; (8) compositions comprising the polypeptide, an agonist
CC compound, an antagonist compound or an antibody, and an excipient; (9)
CC treating diseases or conditions associated with decreased expression or
CC overexpression of functional human PMMM; (10) screening for a compound
CC that is effective as an agonist or antagonist of (1), that specifically
CC binds to (1), that modulates the activity of (1), or is effective in
CC altering expression of the target polynucleotide; (11) screening for a
CC potential toxicity of a test compound; (12) a diagnostic test for a
CC condition or disease associated with the expression of PMMM in a
CC biological sample; (13) diagnosing a condition or disease associated with
CC the expression of PMMM in a subject; (14) preparing a polyclonal or
CC monoclonal antibody with the specificity of the antibody in (6); (15) a
CC polyclonal or monoclonal antibody produced by the method in (14); (16)
CC compositions comprising the polyclonal or monoclonal antibody, and a
CC carrier; (17) generating an expression profile of a sample containing the
CC polynucleotides; and (18) an array comprising different nucleotide
CC molecules affixed at distinct physical locations on a solid substrate,
CC where at least one nucleotide molecule comprises a first oligonucleotide
CC or polynucleotide sequence specifically hybridisable with at least 30
CC contiguous nucleotides of the target polynucleotide. PMMM sequences have
CC gastrointestinal, cardiovascular, immunosuppressive, antiinflammatory,
CC cytostatic, neuroprotective and gynaecological activities, and can be
CC used in gene therapy. The composition and methods are useful for
CC diagnosing, preventing or treating diseases or conditions associated with
CC aberrant expression of PMMM, such as gastrointestinal, cardiovascular,
CC autoimmune/inflammatory, cell proliferative, developmental, epithelial,
CC neurological or reproductive disorders. They may also be used for
CC assessing the effects of exogenous compounds on the expression of nucleic
CC acid and amino acid sequences of PMMM. The PMMM or its fragments are also
CC useful in screening compounds for effectiveness as agonist or antagonist
CC of the polypeptides, or in altering the expression of the target
CC polynucleotide and compounds that specifically bind to or modulate the
CC activity of the polypeptide. The microarray is useful in monitoring or
CC measuring protein-protein interactions, drug-target interactions, and
CC gene expression profiles.
XX
SQ Sequence 498 AA;
Query Match 84.6%; Score 2503; DB 8; Length 498;
Best Local Similarity 86.8%; Pred. No. 7.5e-251;
Matches 495; Conservative 2; Mismatches 1; Indels 72; Gaps 1;
Qy 1 MSRLKAKPSLCKSEPLTTERVTTLSVLKRVTSYCGPSGRLLKQHNFGGVVCTTSQS 60
Db 1 MSRLKAKPSLCKSEPLTTERVTTLSVLKRVTSYCGPSGRLLKQHNFGGVVCTTSQS 60
Qy 61 SALLSHLLVTHPILKILTASI QNHVSSFSDCGLFTAILCCNLINENQVRLGLPTTVIRLN 120
Db 61 SALLSHLLVTHPILKILTASI QNHVSSFSDCGLFTAILCCNLINENQVRLGLPTTVIRLN 120
Qy 121 KHLISLCISYLSKSDTCGCRIPVDVFSSTQILLCLVRSILTSKPACMLTRKETEHSALILR 180
Db 121 KHLISLCISYLSKSDTCGCRIPVDVFSSTQILLCLVRSILTSKPACMLTRKETEHSALILR 180
Qy 181 AFLLTIPENAEGHIIIGKSLIVPLKQORVDSVTLPGLIEMSEVOLMELLPIKKTALK 240
Db 181 AFLLTIPENAEGHIIIGKSLIVPLKQORVDSVTLPGLIEMSEVOLMELLPIKKTALK 240
Qy 241 VALFCTTLTGSDTSGTGVVSVGVLENVAVDQLNLGRQLISDHVDLVLQKVIHPS 300
Db 241 VALFCTTLTGSDTSGTGVVSVGVLENVAVDQLNLGRQLISDHVDLVLQKVIHPS 300
Qy 301 LKQFLNMHRIIAIDRIGVTLMPELTKMTGTQPIGSLGSI CPNSVSGVKDVTAKFGSOHF 360
Db 301 LKQFLNMHRIIAIDRIGVTLMPELTKMTGTQPIGSLGSI CPNSVSGVKDVTAKFGSKHF 360

QY 361 FHILPNEATCSLLLCNRNDTAWDELKLTCTOTALHVLQTLTKEPWALLGGGCTETHLAA 420
Db 361 FHILPNEATCSLLLCNRNDTAWDELKLTCTOTALHVLQTLTKEPWALLGGGCTETHLAA 420
QY 421 IRHKTHNDPESILKDDDECTQTELQLIAEAFCSALESVVGSLHDGGGILLTDMKYGHLWSV 480
Db 421 IRHK----- 424
QY 481 QADSPCVANWPDLLSCGCGLYNSQBELNWSFLRTRRPVPOSCLPHEAVGASNLITLD 540
Db 425 -----CGCGLYNSQBELNWSFLRTRRPVPOSCLPHEAVGASNLITLD 468
QY 541 CLTAKLSGLQVAVETANLIWDLVSVIDKN 570
Db 469 CLTAKLSGLQVAVETANLIWDLVSVIDKN 498

RESULT 6
ABG19636
ID ABG19636 standard; protein; 474 AA.
XX
AC ABG19636;
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #19627.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS83823.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 49995; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (I). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this

CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 474 AA;
Query Match 78.6%; Score 2325; DB 4; Length 474;
Best Local Similarity 97.2%; Pred. No. 2.3e-232;
Matches 458; Conservative 4; Mismatches 7; Indels 2; Gaps 2;
QY 1 MSRLRAKPSLCKSEPLTTRVVRTTILSVLKRIVTSCYGPSGRLLKQLHNGFGGVCTTSQS 60
Db 3 MSRLRAKPSLCKSEPLTTRVVRTTILSVLKRIVTSCYGPSGRLLKQLHNGFGGVCTTSQS 62
QY 61 SALLSHLLVTHPILKILTASIQNHVSSFSDCGLFTAILCCNLIENTVQRLGLPTTVIRLN 120
Db 63 SALLSHLLVTHPILKILTASIQNHVSSFSDCGLFTAILCCNLIENTVQRLGLPTTVIRLN 122
QY 121 KHLISLCISYLSKSDTCGCRIPVDFSGSTOILLCLVRSILTSKPACMLTRKETHSVSAIILR 180
Db 123 KHLISLCISYLSKSETCGCRIPVDFSGSTOILLCLVRSILTSKPACMLTRKETHSVSAIILR 182
QY 181 AFLTIPENAEGHIIILGKSLIVPLKGQORVIDSTVLPGLIEMSEVOLMRLPIKSTALK 240
Db 183 AFLTIPENAEGHIIILGKSLIVPLKGQORVIDSTVLPGLIEMSEVOLMRLPIKSTALK 242
QY 241 VALFCTTSLGSDTSGEGTVVSVGSLENVAVDQLNLNLRQLISDHVDLVCQKVIHPS 300
Db 243 VALFCTTSLGSDTSXPGEGTVVSVGSLENVAVDQLNLNLRQLISDHVDLVCQKVIHPS 302
QY 301 LKQFLNMHPHIIAIDRIGV-TLMEPL-TKMTGTQPTGSLGSI CPNSVGSVKDVCTAKFGSQ 358
Db 303 LKQFLNMHPHIIAIDRIGSDSDVEPLDKWTGTQPTGSLGSI CPNSVGSVKDVCTAKFGSK 362
QY 359 HFFHLLIPNEATCSLLLCNRNDTAWDELKLTCTOTALHVLQTLTKEPWALLGGGCTETHLA 418
Db 363 HFFHLLIPNEATCSLLLCNRNDTAWDELKLTCTOTALHVLQTLTKEPWALLGGGCTETHLA 422
QY 419 AYIRKTHNDPESILKDDDECTQTELQLIAEAFCSALESVVGSLHDGGGILL 469
Db 423 AYIRKTHNDPESILKDDDECTQTELQLIAEAFCSALESVVGSLHDGGGILL 473

RESULT 7
AAB43205
ID AAB43205 standard; protein; 406 AA.
XX
AC AAB43205;
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2969 polypeptide sequence SEQ ID NO:5938.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunosuppressant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disease; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.

```
XX PF 31-MAR-2000; 2000WO-US008621.
XX PR 02-APR-1999; 99US-0127607P.
XX PR 02-APR-1999; 99US-0127636P.
XX PR 03-APR-1999; 99US-0127728P.
XX PR 30-MAR-2000; 2000US-00540763.
XX PA (CURA-) CURAGEN CORP.
XX PI Shimkets RA, Leach M;
XX DR WPI; 2000-602362/57.
XX DR N-PSDB; AAC77414.
XX PT Novel nucleic acids and peptides derived from open reading frame X,
XX PT useful for treating e.g. cancers, proliferative disorders,
XX PT neurodegenerative disorders and cardiovascular disease.
XX PS Claim 11; Page 5122-5123; 5507pp; English.
XX CC AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
XX CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX CC antipsoriatic; antiparkinsonian; nontropic; neuroprotective; osteopathic;
XX CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
XX CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
XX CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
XX CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
XX CC sequences can be used for determining the presence of or predisposition
XX CC to, or preventing or treating pathological conditions associated with an
XX CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
XX CC proteins in gene therapy vectors. The proteins and nucleic acids may be
XX CC used to treat cancers, proliferative disorders, neurodegenerative
XX CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
XX CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
XX CC storage, systemic lupus erythematosus, severe combined immunodeficiency
XX CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
XX CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
XX CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
XX CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX SQ Sequence 406 AA;
Query Match 70.5%; Score 2086; DB 3; Length 406;
Best Local Similarity 99.0%; Pred. No. 1.4e-207;
Matches 402; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 165 MLTRKETEHSVALILRAFLITIPENAEHIIIGKSLIVPLKQGVVIDSTVLPGLIEMSE 224
DB 1 MLTRKETEHSVALILRAFLITIPENAEHIIIGKSLIVPLKQGVVIDSTVLPGLIEMSE 60
QY 225 VOLMLLPIKSTALKVALFCTTLGSDTSGTGTVVSVGSLENVADQLNLGRQLI 284
DB 61 VOLMLLPIKSTALKVALFCTTLGSDTSGTGTVVSVGSLENVADQLNLGRQLI 120
QY 285 SDHVDLVLCQVTHPSLKQFLNMRHIIADRIQVTLMEPLTKMTGTQPTGSGISICPNY 344
DB 121 SDHVDLVLCQVTHPSLKQFLNMRHIIADRIQVTLMEPLTKMTGTQPTGSGISICPNY 180
QY 345 GSVKDVCTAFKGSQHPFLHIPNEATICSLLNCRNDTAWDELKLTCTQALHVLQTLKEP 404
DB 181 GSVKDVCTAFKGSQHPFLHIPNEATICSLLNCRNDTAWDELKLTCTQALHVLQTLKEP 240
QY 405 WALLGGCGTETHLAAYIRKHTNDPESILKDDCTOTELQLIAEAPCSALESVVGSLHED 464
DB 241 WALLGGCGTETHLAAYIRKHTNDPESILKDDCTOTELQLIAEAPCSALESVVGSLHED 300
QY 465 GGBILDMKYGHLSVQADSPCVANWPDLLSQCGLYNSQBELNWSFLRSTRPVPQS 524
DB 301 GGBILDMKYGHLSVQADSPCVANWPDLLSQCGLYNSQBELNWSFLRSTRPVPQS 360
QY 525 CLPHEAVGSASNLTLDCLTAKLSGLQVAVETANLIWDLGVVIEDKN 570
```

```
DB 361 CLPHEAVGSASNLTLDCLTAKLSGLQVAVETANLIWDLGVVIEDKN 406
RESULT 8
AAB56236
ID AAB56236 standard; protein; 222 AA.
XX AC AAB56236;
XX DT 13-MAR-2001 (first entry)
XX DE Human secreted protein sequence encoded by gene 142 SEQ ID NO:330.
XX KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
XX KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
XX KW cerebroprotective; nontropic; neuroprotective; antibacterial; virucide;
XX KW fungicide; ophthalmological; gene therapy; pathological condition;
XX KW autoimmune disease; rheumatoid arthritis; hyperproliferative ischaemia;
XX KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
XX KW cerebrovascular disorder; angiogenesis; nervous system disorder;
XX KW Alzheimer's disease; infection; ocular disorder; corneal infection;
XX KW wound healing; skin aging; food additive; preservative; chromosome 20.
XX OS Homo sapiens.
XX PN WO200070042-A1.
XX PD 23-NOV-2000.
XX PF 11-MAY-2000; 2000WO-US012788.
XX PR 13-MAY-1999; 99US-0134068P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE,
XX PI Duan RD, Florence KA, Soppet DR;
XX XX WPI; 2000-679828/66.
XX DR N-PSDB; AAC99977.
XX PT Isolated nucleic acid molecule encoding a human secreted protein is used
XX PT in preventing, treating or ameliorating a medical condition.
XX PS Claim 11; Page 1007; 1065pp; English.
XX CC The polynucleotide sequences given in AAC99818 to AAC99977 encode the
XX CC human secreted proteins given in AAB56077 to AAB56362. Human secreted
XX CC proteins have activities based on the tissues and cells the genes are
XX CC expressed in. Examples of activities include: immunosuppressive;
XX CC antarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
XX CC vasotropic; cerebroprotective; nontropic; neuroprotective; antibacterial;
XX CC virucide; fungicide; and ophthalmological. The human secreted
XX CC polynucleotides and proteins can be used to prevent, treat or ameliorate
XX CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
XX CC dogs, chickens or sheep. They are also used in diagnosing a pathological
XX CC condition or susceptibility to a pathological condition. Disorders which
XX CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid
XX CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
XX CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
XX CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
XX CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
XX CC fungi and ocular disorders e.g. corneal infection. The proteins can also
XX CC be used to aid wound healing and epithelial cell proliferation, to
XX CC prevent skin aging due to sunburn, to maintain organs before
XX CC transplantation, for supporting cell culture of primary tissues, to
XX CC regenerate tissues and in chemotaxis. The proteins can also be used as a
XX CC food additive or preservative to increase or decrease storage
XX CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
XX CC in the exemplification of the present invention
XX SQ Sequence 222 AA;
```

Query Match 36.9%; Score 1092; DB 3; Length 222;
Best Local Similarity 97.6%; Pred. No. 1.6e-104;
Matches 206; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 360 FFLHPIEATICSLLLCNRNDTAWDELKLTCTQTALHVLQTLKEPWALLGGGCTETHLAA 419
DB 12 FFLHPIEATICSLLLCNRNDTAWDELKLTCTQTALHVLQTLKEPWALLGGGCTETHLAA 71

QY 420 YIRHKTHNDPESILKDDCTQTQLIAEAFCSALESVVSGSLRHDGGEILTDMKYGHLMS 479
DB 72 YIRHKTHNDPESILKDDCTQTQLIAEAFCSALESVVSGSLRHDGGEILTDMKYGHLMS 131

QY 480 VQADSCFVANWPDLLSQCGGLYNSQBELNWSFLRSTRFPVQSCLPHEAVGSASNLTL 539
DB 132 VQADSCFVANWPDLLSQCGGLYNSQBELNWSFLRSTRFPVQSCLPHEAVGSASNLTL 191

QY 540 DCLTAKLSGLQAVETANLILDSYVIEDKN 570
DB 192 DCLTAKLSGLQAVETANLILDSYVIEDKN 222

RESULT 9
AAB56218
ID AAB56218 standard; protein; 223 AA.
XX AAB56218;
XX
XX 13-MAR-2001 (first entry)
XX Human secreted protein sequence encoded by gene 142 SEQ ID NO:312.
XX
XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; neutropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; gene therapy; pathological condition;
KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
KW cerebrovascular disorder; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; skin aging; food additive; preservative; chromosome 20.
XX
OS Homo sapiens.
XX
XX WO200070042-A1.
XX
XX 23-NOV-2000.
XX
XX 11-MAY-2000; 2000WO-US012788.
XX
XX 13-MAY-1999; 99US-0134068P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CB;
PI Duan RD, Florence KA, Soppet DR;
XX
XX WPI; 2000-679828/66.
DR N-PSDB; AAC99959.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is used
PT in preventing, treating or ameliorating a medical condition.
XX
XX Claim 11; Page 999-1000; 1065pp; English.
XX
XX The polynucleotide sequences given in AAC99818 to AAC99977 encode the
CC human secreted proteins given in AAB56077 to AAB56362. Human secreted
CC proteins have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: immunosuppressive;
CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
CC vasotropic; cerebroprotective; neutropic; neuroprotective; antibacterial;
CC virucide; fungicide; and ophthalmological. The human secreted
CC polynucleotides and proteins can be used to prevent, treat or ameliorate

CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
CC dogs, chickens or sheep. They are also used in diagnosing a pathological
CC condition or susceptibility to a pathological condition. Disorders which
CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid
CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
CC fungi and ocular disorders e.g. corneal infection. The proteins can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The proteins can also be used as a
CC food additive or preservative to increase or decrease storage
CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
CC in the exemplification of the present invention
XX
SQ Sequence 223 AA;
Query Match 36.4%; Score 1076; DB 3; Length 223;
Best Local Similarity 98.6%; Pred. No. 7.4e-103;
Matches 213; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 165 MLTRKETEHSVALLIRAFLLTIPENAEHIIILGKSLIVPLKQGVVDSTVLPGLIEMSE 224
DB 1 MLTRKETEHSVALLIRAFLLTIPENAEHIIILGKSLIVPLKQGVVDSTVLPGLIEMSE 60

QY 225 VQLMRLLPIKKSSTALKVALFCTTLSDGTSDTGEGTVVSGVSLNNAVLDQLNLGRQLI 284
DB 61 VQLMRLLPIKKSSTALKVALFCTTLSDGTSDTGEGTVVSGVSLNNAVLDQLNLGRQLI 120

QY 285 SHQVDVLVLCQVTHPSLKQFLANHRIIADRGVTLMEPLTKMTGTQPTGSLGSI CPNSY 344
DB 121 SHQVDVLVLCQVTHPSLKQFLANHRIIADRGVTLMEPLTKMTGTQPTGSLGSI CPNSY 180

QY 345 GSVKDVCTAKFGSQHFFHILPNEATICSLLLCNRND 380
DB 181 GSVKDVCTAKFGSKHFFHILPNEATICSLLLCNRNE 216

RESULT 10
AAB56356
ID AAB56356 standard; protein; 166 AA.
XX AAB56356;
XX
XX 13-MAR-2001 (first entry)
XX Human secreted protein sequence encoded by gene 142 SEQ ID NO:450.
XX
XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; neutropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; gene therapy; pathological condition;
KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
KW cerebrovascular disorder; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; skin aging; food additive; preservative; chromosome 20.
XX
OS Homo sapiens.
XX
XX WO200070042-A1.
XX
XX 23-NOV-2000.
XX
XX 11-MAY-2000; 2000WO-US012788.
XX
XX 13-MAY-1999; 99US-0134068P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CB;
PI

PI Duan RD, Florence KA, Soppet DR;
XX WPI; 2000-679828/66.
XX
PT Isolated nucleic acid molecule encoding a human secreted protein is used
PT in preventing, treating or ameliorating a medical condition.
XX
PS Disclosure; Page 1060; 1065pp; English.
XX
XX The polynucleotide sequences given in AAC99818 to AAC99977 encode the
CC human secreted proteins given in AAB56077 to AAB56362. Human secreted
CC proteins have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: immunosuppressive;
CC antithrombotic; antirheumatic; antiproliferative; cytostatic; cardiant;
CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
CC virucide; fungicide; and ophthalmological. The human secreted
CC polynucleotides and proteins can be used to prevent, treat or ameliorate
CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
CC dogs, chickens or sheep. They are also used in diagnosing a pathological
CC condition or susceptibility to a pathological condition. Disorders which
CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid
CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
CC fungi and ocular disorders e.g. corneal infection. The proteins can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The proteins can also be used as a
CC food additive or preservative to increase or decrease storage
CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
CC in the exemplification of the present invention
XX
SQ Sequence 166 AA;

Query Match 28.4%; Score 841; DB 3; Length 166;
Best Local Similarity 99.4%; Pred. No. 1.4e-78;
Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRLKAKPSLCKSEPLTTERVTTLSVLKRIVTCYGPSGRLLKQLHNGFGGVCTTSQS 60
DB 3 MSRLKAKPSLCKSEPLTTERVTTLSVLKRIVTCYGPSGRLLKQLHNGFGGVCTTSQS 62
QY 61 SALLSHLLVTHPIKILTASIQNHVSFSDCGLFTALCCNLLENVORLGLTPTTVIRLN 120
DB 63 SALLSHLLVTHPIKILTASIQNHVSFSDCGLFTALCCNLLENVORLGLTPTTVIRLN 122
QY 121 KHLLSLCISYLSKSDTCGRIPVDFPSSTQILLCLVRSILTSKPAC 164
DB 123 KHLLSLCISYLSKSDTCGRIPVDFPSSTQILLCLVRSILTSKPAC 166

RESULT 11
AAG02893
ID AAG02893 standard; protein; 86 AA.
XX
AC AAG02893;
XX
XX
DT 06-OCT-2000 (first entry)
XX
XX Human secreted protein, SEQ ID NO: 6974.
XX
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
XX Homo sapiens.
XX
PN EF1033401-A2.
XX
XX
PD 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-00200610.

XX
PR 26-FEB-1999; 99US-0122487P.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duciart A, Giordano J;
XX
XX WPI; 2000-500381/45.
DR N-PSDB; AAC02899.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
XX Claim 13; SEQ ID NO 6974; 71pp + Sequence Listing; English.
XX
XX The present sequence is a polypeptide encoded by one of a large number of
CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
CC prepared from total human RNAs or polyA+ RNAs derived from 30 different
CC tissues. EST sequences usually correspond mainly to the 3' untranslated
CC region (UTR) of the mRNA because they are often obtained from oligo-dT
CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
CC sequences derived from the 5' ends of mRNAs and even in those cases where
CC longer cDNA sequences have been obtained, the full 5' UTR is rarely
CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can
CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
CC are also used in diagnostic, forensic, gene therapy and chromosome
CC mapping procedures. They are used to obtain upstream regulatory sequences
CC and to design expression and secretion vectors
XX
SQ Sequence 86 AA;

Query Match 14.6%; Score 432; DB 3; Length 86;
Best Local Similarity 98.8%; Pred. No. 1.8e-36;
Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSRLKAKPSLCKSEPLTTERVTTLSVLKRIVTCYGPSGRLLKQLHNGFGGVCTTSQS 60
DB 1 MSRLKAKPSLCKSEPLTTERVTTLSVLKRIVTCYGPSGRLLKQLHNGFGGVCTTSQS 60
QY 61 SALLSHLLVTHPIKILTASIQNHVS 86
DB 61 SALLSHLLVTHPIKILTASIQNHVS 86

RESULT 12
ADN18515
ID ADN18515 standard; protein; 538 AA.
XX
AC ADN18515;
XX
XX 02-DEC-2004 (first entry)
XX
XX Bacterial polypeptide #1168.
XX
XX Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX
XX Bacteria.
XX
XX US2003233675-A1.
XX
XX 18-DEC-2003.
XX
XX 20-FEB-2003; 2003US-00369493.
XX
XX 21-FEB-2002; 2002US-0360039P.
XX

PA (CAOX/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 1168; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 538 AA;

Query Match 8.7%; Score 257.5; DB 8; Length 538;
Best Local Similarity 24.1%; Pred. No. 4.8e-17;
Matches 115; Conservative 84; Mismatches 214; Indels 65; Gaps 19;
QY 17 LTTERTVTLVSLKRIVTCVGPGRKQLHNGFGVGVCTTSQSALLSHLLVTHPIKI 76
DB 33 ILAETVRTL-----GPKGMDKMLVDSLGDIV-VTNDGVILKEMDIEHPAAKM 80
QY 77 LTASIQNHVSSFDGGLFTAILCCNLI-----ENVQRILGTPPTVIRLNKHLSLCISYLK 132
DB 81 LVEVAKTQDEVDG-GTTTAVIAGELLKKAENLEIHP-TIIAMGYRQAQAEQEL 138
QY 133 SDTCGGRIPVDFSTQILLCLVRSILTSKPCMLTRKEHVSALLIRAPFLITPENAE- 191
DB 139 DD-----IAIDASDRDTLMKVAMTMTGKG----TEKAREPLAELIVDA-VKQVEEDGEV 188
QY 192 --GHILGKSLIVPLKGORVIDSTVPLGILIEMSEVQLMRLPIKSKTALVFCVTLIS 249
DB 189 EKDHIEKK-----EGAAVDDSTLVQGVIIKERVH--PGMP-KKVENNKIALLNCPI- 239
QY 250 GDTSDTGEGTVVSYGVSVLENALVDQLNLGROLISDHVD-----LVLCQKRVHPSLKQFL 305
DB 240 -EVKETEVDAREITDPSQMAFTEQEQMIRDMVNSIVDTGANVLFQCKGIDDLAQHYL 298
QY 306 NMRIITADIRGVTLMETLMTGTQPIGSLGICPNSYGVKDVCTAKP-GSQHFHIL 364
DB 299 AKAGVLAVRVKXSDMEKLSKATGANIVTNIEDLSPEDLGAGVWSEKISGSEMFVEE 358
QY 365 PNEATICSLLLCNRNDTAWDELKLTCTALHVLQLTLPKFWALLGGCGCTETHLAAVIRHK 424

DB 359 CKPEKAVTILVRGSTEHVHSEVERAIEDAIGVVAATVEDGKVVAGGAPAEIEA----- 412
QY 425 THNDPESILKD--DECTQTEFQLIABAFCSALESVGSLEHDPG-----BILDMKYGH 476
DB 413 -----KRLKDYADSIISGRE-QLAVSAFAEALEIVPKTLAENAGLSDISDLVLDLRAH 463
RESULT 13
AAO24120
XX AAO24120 standard; protein; 560 AA.
XX
XX AAO24120;
XX 11-MAR-2004 (first entry)
XX
XX S shibatae TF55 alpha subunit.
XX
XX Chaperonin; nanotemplate; nanostructure; nanoarray; nanodevice; TF55;
KW human; mouse; yeast.
XX
XX Sulfolobus shibatae.
XX
XX WO2003080796-A2.
XX
XX 02-OCT-2003.
XX
XX 08-NOV-2002; 2002WO-US035889.
XX
XX 08-NOV-2001; 2001US-0340538P.
XX
XX (USAS) NASA US NAT AERO & SPACE ADMIN.
XX (SETI-) SETI INST.
XX
XX Trent JD, Mcmillan RA, Kagawa H, Paavola CD;
XX WPI; 2003-845205/78.
XX
XX Novel chaperonin polypeptides, useful in field-effect transistor,
PT integrated circuit, field emitter, probe, nanocomputer, quantum computer,
PT sensor or biochip applications.
XX
XX Disclosure; Fig 15; 91pp; English.
XX
XX The present invention relates to an isolated polypeptide comprising an S.
CC shibatae TF55 beta subunit and its mutants. Also described are
CC nanotemplates and devices comprising at least one chaperonin protein such
CC as TF55. A nanodevice comprising at least one nanotemplate is used as a
CC field-effect transistor, integrated circuit, field emitter, probe,
CC nanocomputer, quantum computer or sensor. The nanodevice is a
CC nanoelectromechanical system. The nanodevice is an optoelectronic switch,
CC optical switch, lens, laser, photonic crystals, or amplifier. The
CC nanodevice is a magnetic memory device, a magnetic storage device, hard-
CC disk drive read heads, magnetic RAM, magnetic field sensor, magnetic
CC logic gates or a magnetic switch. The nanodevice is an electronic,
CC semiconductor, mechanical, electromechanical, magnetic, photonic,
CC optical, or optoelectronic device. The nanoarray is a multi-nanowell
CC assay plate, or a single-molecule probe for DNA detection and
CC hybridisation. The present sequence is a protein shown in the
CC exemplification of the invention
XX
SQ Sequence 560 AA;
Query Match 8.7%; Score 256.5; DB 7; Length 560;
Best Local Similarity 24.3%; Pred. No. 6.5e-17;
Matches 115; Conservative 84; Mismatches 221; Indels 53; Gaps 16;
QY 29 LKRIVTCVGPGRKQLHNGFGVGVCTTSQSALLSHLLVTHPIKILTASTQNHSVSF 88
DB 31 LAEMRSSGPGKGLDKMLDSFGD-VTINDGATIVKMEIQHPAAKLIVKAAQDAEV 89
QY 89 SDGGLFTAILCCNLIENVQRL---GLTPTTVIRLNKHLSLCISYLSKSTCGCRIPVDFS 145
DB 90 GDGTTSAVVLAGALLEKAESLLDNQIHPTIIIEGYKAVTKALELLPQ--LGTRIDIRDL 147

PA (CAOV/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
XX WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
XX
PS Claim 1; SEQ ID NO 22860; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 557 AA;

Query Match 8.4%; Score 248.5; DB 8; Length 557;
Best Local Similarity 23.6%; Pred. NO. 4.4e-16;
Matches 108; Conservative 85; Mismatches 214; Indels 51; Gaps 14;
QY 28 VLKRVITSCYGPGRGLKQLHNGFGVYCTTSQSSALLSHLVTHPILKIITASIQNHVSS 87
DB 36 VLAEMLKSSLGPRGLDKMLVDAFGD-ITVNDGATIVKEMEIOHPAKLLVEVAKQADAE 94
QY 88 FSDCGLFTALCNLIENVQRL---GLTPTTVIRLNKHLISLCISYLSKSDTCGCR----- 139
DB 95 VGDGTTVVVLGALLEKAELDENLHPTIIIE-----GYTKAMEEARLVDEA 144
QY 140 -IPVDFSSTOILCLVRSILTSKPAKMLRKTEHVSALILRAFLITIPENAEQ--HIIIL 196
DB 145 AVPEVEDEDDSVLRRIAEFTTLASK--FVGTGPERDKIISMVIDAIRTVAEKRPDGGYVDL 202
QY 197 GKSLIVPLKQORVIDSVLPGLIEMSEVQLMRLPLPKSTALKVALFCTTISGDTSDTG 256
DB 203 DYVKIEKKGGSLDSKLVKRGIVLDKEVH--PAMP-KRVENAKIIVLDAPLEVQKPEL- 258
QY 257 EGTVVVSYGVSLENVLDQ---LLNLGRQLISDHVDLVCQKVIHPSLKQFLNMHRIIAI 313
DB 259 TTKIRVTDIEKLESFLEETEERMLRDMVEKIAATGANVVIQKGIIDEVAQHFLAKKGILAV 318
QY 314 DRIGVTLMELPTKMTGTQPIGSLGSCPNISYSGVKDVCYAKFGSQH--PFHLIPNEATIC 371
DB 319 RRVKRSIDIEKVAKATGAKIVTSLRDLKPEYLGVAELVEERKVGEDKRVFTGAKNPKSV- 377
QY 372 SLLLCNRNDTAWDELKLTCTQALHLVQLTLKEFPWALIGGGCTETHLAAYIRHKHTNDPES 431

Db 378 TILLRGANDMLDDEARNIKDALHGLRNILREPKIVGGGAVEVEIA----- 424
QY 432 ILKDDECTQT----ELQLIABAFCSALESVVGSLEHDDGG 466
DB 425 -LKLKEFARTVGGKQQLAIEAYAEALETIPTVLAESAG 461
Search completed: September 3, 2005, 04:59:18
Job time : 173 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 09:34:10 ; Search time 7919 Seconds
(without alignments)
11882.166 Million cell updates/sec

Title: US-10-616-263-29

Perfect score: 2472

Sequence: 1 tatgagccttcggaactgtg.....aaaaaaaaaaaaaaaaaaaaa 2472

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gss1:*

9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2361.2	95.5	2481	3	CR597193 full-length
2	1544.2	62.5	2658	3	AK032528 Mus muscu
3	1544.2	62.5	2658	3	AK032554 Mus muscu
4	1198.2	48.5	2745	3	AK004987 Mus muscu
5	880.8	35.6	1105	4	BM473886 AGENCOURT
6	870.4	35.2	1019	4	BM558221 AGENCOURT
7	849.8	34.4	920	5	BQ641810 AGENCOURT
8	849.4	34.4	1137	1	AL532426 AL532426
9	843.4	34.1	930	5	BQ675098 AGENCOURT
10	797.8	32.3	932	5	BUS26879 AGENCOURT
11	785.4	31.8	810	4	BG742334 602631495
12	781	31.6	1044	5	BM913546 AGENCOURT
13	765	30.9	912	5	BQ423217 AGENCOURT
14	752.2	30.4	942	5	BX433357 BX433357
15	738	29.9	1013	1	AL532425 AL532425
16	729.4	29.5	737	4	BI085537 602870248
17	694.8	28.1	926	4	BG397725 602438786
18	693.8	28.1	697	5	BQ574234 UI-H-EZ1-
19	692.4	28.0	710	7	CN428047 170004243
20	674.2	27.3	879	7	CK467658 938975 WA
21	668.4	27.0	696	4	BG749677 602707221
22	663.4	26.8	929	5	BX452469 BX452469
23	660	26.7	722	2	BE279744 601157720
24	658.4	26.6	661	6	CA423280 UI-H-EZ1-

25	656.2	26.5	764	4	BI549136
26	652.6	26.4	693	4	BI767302
27	648.8	26.2	652	5	BU686127
28	642.6	26.0	822	4	BM040705
29	642.4	26.0	644	5	BU947652
30	641.8	26.0	803	7	CR753485
31	638.2	25.8	887	4	BG036503
32	637.4	25.8	639	6	CB141647
33	635.4	25.7	637	4	BM772823
34	634.4	25.7	636	4	BM772193
35	634.4	25.7	636	4	BM772822
36	626.2	25.3	782	4	BI086889
37	623.8	25.2	627	5	BQ549598
38	623.4	25.2	760	2	BE386143
39	620.4	25.1	762	7	CR764167
40	618.4	25.0	691	2	AW967111
41	617.8	25.0	626	5	BM726839
42	617.4	25.0	628	5	BU741135
43	616.4	24.9	640	5	BX093992
44	612.4	24.8	625	4	BG182054
45	606.4	24.5	626	4	BG192494

ALIGNMENTS

RESULT 1	CR597193	2481 bp	mRNA	linear	HTC 21-JUL-2004
LOCUS	full-length cDNA clone CS0DM014YF11 of Fetal liver of Homo sapiens (human).				
DEFINITION	CR597193				
ACCESSION	CR597193.1	GI:50478000			
VERSION	HTC; CINSIT_CDNA.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 2481)				
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.				
TITLE	Full-length cDNA libraries and normalization				
JOURNAL	Unpublished				
REMARK	Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Paraday Avenue				
REFERENCE	2 (bases 1 to 2481)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)				
COMMENT	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.				
FEATURES	Location/Qualifiers				
source	1..2481				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="CS0DM014YF11"				
	/tissue_type="Fetal liver"				
	/plasmid="pCMVSPORT_6"				

ORIGIN	Query Match 95.5%; Score 2361.2; DB 3; Length 2481;
	Best Local Similarity 99.9%; Pred. No. 0;
	Matches 2363; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	1	TATGAGCCTTCGGAACCTGTGGAGAGACTACAAAGTTTGGTTGTTATGTCCTTTAGT	60
DB	116	TATGAGCCTTCGGAACCTGTGGAGAGACTACAAAGTTTGGTTGTTATGTCCTTTAGT	175

QY	61	TGGGCTCATACATTTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTGTTTTCCAAATACCTAA	120
Db	176	TGGGCTCATACATTTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTGTTTTCCAAATACCTAA	235
QY	121	AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTTCAAAATCTTTAGAGAGGCGCAAAAT	180
Db	236	AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTTCAAAATCTTTAGAGAGGCGCAAAAT	295
QY	181	CCAGGGGAAGTAGCAGGCTTGCAATCTTCAGGTAAAGAGCAGCTTTTGAATCTGAGCTTC	240
Db	296	CCAGGGGAAGTAGCAGGCTTGCAATCTTCAGGTAAAGAGCAGCTTTTGAATCTGAGCTTC	355
QY	241	ATATCCAAAGAGAGATGAAATATACAGTTGGATTAGAAAGACTGGCTTCTTGTAGCT	300
Db	356	ATATCCAAAGAGAGATGAAATATACAGTTGGATTAGAAAGACTGGCTTCTTGTAGCT	415
QY	301	GGGATATCTTTTCATAGGTGTCCATCTTTGGAAACATACCTTTTGCAGAGGTCTGCAAAAGCAG	360
Db	416	GGGATATCTTTTCATAGGTGTCCATCTTTGGAAACATACCTTTTGCAGAGGTCTGCAAAAGCAG	475
QY	361	TCGTAAAAATTCAGTCTCAAAGCAAAACAAAAGAGATTGAAGAGTGAAGTAAAAATTAAT	420
Db	476	TCGTAAAAATTCAGTCTCAAAGCAAAACAAAAGAGATTGAAGAGTGAAGTAAAAATTAAT	535
QY	421	ATTTGGAATTTACTAATTTGTTCATTAATCATTTCTATGCTGATTAGCTTCATAAAACATTGA	480
Db	536	ATTTGGAATTTACTAATTTGTTCATTAATCATTTCTATGCTGATTAGCTTCATAAAACATTGA	595
QY	481	ACTTTTTGATTTTTATAGCCCAATGCTGCATATTTTCATCTTTAAATTCCTAAAGAAATTAAT	540
Db	596	ACTTTTTGATTTTTATAGCCCAATGCTGCATATTTTCATCTTTAAATTCCTAAAGAAATTAAT	655
QY	541	TTTAAATGTTAAACGTTGATTAATGCAATAAATAGAAAAATGTGGTTTACAAAAATAAAACG	600
Db	656	TTTAAATGTTAAACGTTGATTAATGCAATAAATAGAAAAATGTGGTTTACAAAAATAAAACG	715
QY	601	GTCTTCACATGTTACACCTGAAGTAAGATGTCTCGTTTGGAGCTTGAAGCCATCATTT	660
Db	716	GTCTTCACATGTTACACCTGAAGTAAGATGTCTCGTTTGGAGCTTGAAGCCATCATTT	775
QY	661	GTGTAAGAGTGAACCACTGACAACTGAGAGAGTCAGGACCAACATTTTCTGTCTTGAAAAAG	720
Db	776	GTGTAAGAGTGAACCACTGACAACTGAGAGAGTCAGGACCAACATTTTCTGTCTTGAAAAAG	835
QY	721	AATTGTAAACATCATGTATGGCCCTCAGGTAGGCTGAAAGCAGCTGCACAATGGCTTTGG	780
Db	836	AATTGTAAACATCATGTATGGCCCTCAGGTAGGCTGAAAGCAGCTGCACAATGGCTTTGG	895
QY	781	AGGTTACGTTGTACAACTCAGTCCTCAGCTCTGCTCAGTCACCTTTTGGTCACACA	840
Db	896	AGGTTACGTTGTACAACTCAGTCCTCAGCTCTGCTCAGTCACCTTTTGGTCACACA	955
QY	841	TCCATTTTAAAGATCCTGACAGCCTCCATACAGAAATCATGTGTCAAGCTTCAGTGATTG	900
Db	956	TCCATTTTAAAGATCCTGACAGCCTCCATACAGAAATCATGTGTCAAGCTTCAGTGATTG	1015
QY	901	TGGCTTATTCACAGCTATTCTTTGTGCAACCTGATTGAAAAATGTTTCAGAGATTAGGCTTT	960
Db	1016	TGGCTTATTCACAGCTATTCTTTGTGCAACCTGATTGAAAAATGTTTCAGAGATTAGGCTTT	1075
QY	961	GACACCCACACCTGCTATAGATTAAATAAACAATCTTTTGGCTCTTTGTCATCAGTTATCTT	1020
Db	1076	GACACCCACACCTGCTATAGATTAAATAAACAATCTTTTGGCTCTTTGTCATCAGTTATCTT	1135
QY	1021	CAAGTCTGATPACCTGTGGTTGTGCAATCCAGGTGGAATTTAGTAGTACTCAGATCCTCCT	1080
Db	1136	CAAGTCTGAGACCTGTGGTTGTGCAATCCAGGTGGAATTTAGTAGTACTCAGATCCTCCT	1195
QY	1081	TTGTTTGGTGGTGTATATTAACAAGTAAACCTGCTGTATGCTCACCAAGAAAGAAAC	1140
Db	1196	TTGTTTGGTGGTGTATATTAACAAGTAAACCTGCTGTATGCTCACCAAGAAAGAAAC	1255

QY	1141	AGAGCATGTCACTGCTTTTGATCTCTGAGAGCCTTTTGTCTTACAAATTCAGAAAAATGCTCGA	1200
Db	1256	AGAGCATGTCACTGCTTTTGATCTCTGAGAGCCTTTTGTCTTACAAATTCAGAAAAATGCTCGA	1315
QY	1201	AGGCCACATCATTTTATAGGAAAGAGTTTAAATTTGTACCTTTTAAAGGTCAAAGAGTTATAGA	1260
Db	1316	AGGCCACATCATTTTATAGGAAAGAGTTTAAATTTGTACCTTTTAAAGGTCAAAGAGTTATAGA	1375
QY	1261	TTTCCATCTGTATTTACTCTGGGATATCTCATTTGAAATGTGAGAGTTCAATTTAATGAGGCTATT	1320
Db	1376	TTTCCATCTGTATTTACTCTGGGATATCTCATTTGAAATGTGAGAGTTCAATTTAATGAGGCTATT	1435
QY	1321	ACCTATCAAAAAATCAACTGCCCCCAAGGTGGACCTCTTTTGTACAACCTTTATCCGGAGA	1380
Db	1436	ACCTATCAAAAAATCAACTGCCCCCAAGGTGGACCTCTTTTGTACAACCTTTATCCGGAGA	1495
QY	1381	CACCTTCTGACACTGGAGAGAGAACTGTGGTGGTCAAGTTATGGGGTTTCTTTGAAAAATGC	1440
Db	1496	CACCTTCTGACACTGGAGAGAGAACTGTGGTGGTCAAGTTATGGGGTTTCTTTGAAAAATGC	1555
QY	1441	AGCTTTGGACCAAGCTGCTTAACTAGGAAGGAGCAGCTAATCAGTGACCACTGATCTTGT	1500
Db	1556	AGCTTTGGACCAAGCTGCTTAACTAGGAAGGAGCAGCTAATCAGTGACCACTGATCTTGT	1615
QY	1501	CCTGTGCCAAAAAGTTATACATCCATCTTTGAAGCAGTTTCTCAATATGTCATGCTATTAT	1560
Db	1616	CCTGTGCCAAAAAGTTATACATCCATCTTTGAAGCAGTTTCTCAATATGTCATGCTATTAT	1675
QY	1561	TGCCATAGACAGAAATTTGGAAGTCACTCTGATGAAACCCCTGACTAAAAATGACAGGAACACA	1620
Db	1676	TGCCATAGACAGAAATTTGGAAGTCACTCTGATGAAACCCCTGACTAAAAATGACAGGAACACA	1735
QY	1621	GCCTATTTGGATCCCTAGGCTCAATATGCTCTAATAGTTATAGGAAGTGTGAAAGAGTGTG	1680
Db	1736	GCCTATTTGGATCCCTAGGCTCAATATGCTCTAATAGTTATAGGAAGTGTGAAAGAGTGTG	1795
QY	1681	CACCTGCAAAATTTGGCTCCCAACATTTTTTTCATCTTATTCCTAATGAAGCAACATCTG	1740
Db	1796	CACCTGCAAAATTTGGCTCCCAACATTTTTTTCATCTTATTCCTAATGAAGCAACATCTG	1855
QY	1741	CAGCTTGTCTTCTCTGCAACAGAAAAATGACACTGCTCCCTGGGATGAGCTGAAGCTCACGTGTCA	1800
Db	1856	CAGCTTGTCTTCTCTGCAACAGAAAAATGACACTGCTCCCTGGGATGAGCTGAAGCTCACGTGTCA	1915
QY	1801	GACGGCACTGCAATGCTCTGCAAGTAACTCAAGGAAACATGGGCTTTTGTGGAGGTGG	1860
Db	1916	GACGGCACTGCAATGCTCTGCAAGTAACTCAAGGAAACATGGGCTTTTGTGGAGGTGG	1975
QY	1861	CTGTACTGAAACCTCATTTTGGCTGCATATATCAGACACAGACTCACAGACCCAGAAAG	1920
Db	1976	CTGTACTGAAACCTCATTTTGGCTGCATATATCAGACACAGACTCACAGACCCAGAAAG	2035
QY	1921	CATTCTCAAAGATGATGATGACTCAAAACAGAACTTCAATTTAAATTTGCTGAAGCAATTTTG	1980
Db	2036	CATTCTCAAAGATGATGATGACTCAAAACAGAACTTCAATTTAAATTTGCTGAAGCAATTTTG	2095
QY	1981	CAGTGGCCCTAGAAATCTGTGTTGGCTCTTTAGAAACATGATGAGAGTGAATTTCTCACTGA	2040
Db	2096	CAGTGGCCCTAGAAATCTGTGTTGGCTCTTTAGAAACATGATGAGAGTGAATTTCTCACTGA	2155
QY	2041	CATGAAGTATGACACCTTTTGGTCAAGTTCAGGACAGATTCCTCCTGTGTTGCTAACTGGCC	2100
Db	2156	CATGAAGTATGACACCTTTTGGTCAAGTTCAGGACAGATTCCTCCTGTGTTGCTAACTGGCC	2215
QY	2101	AGATTGCTTTTCAAGTGTGGCTGTGGATTATACAATAGCCAGGAAGAACTCAACTGGTTC	2160
Db	2216	AGATTGCTTTTCAAGTGTGGCTGTGGATTATACAATAGCCAGGAAGAACTCAACTGGTTC	2275
QY	2161	TTTCTTAAAGAACACACGTGCTCCATTTGTGCGCAACAAAGCTGCTTCCACATGAAGCTGT	2220
Db	2276	TTTCTTAAAGAACACACGTGCTCCATTTGTGCGCAACAAAGCTGCTTCCACATGAAGCTGT	2335
QY	2221	GGGCTCAGCCAGCAACCTGACCTTGGACTGTTTGTGACTGCAAAAGCTTATGGGCTACAGGT	2280

```

Db      2336 GGGCTCAGCAGCAACCTGACCTGGACCTGTTTCACTGCAAGCTTAGTGCGCTACAGGT 2395
Qy      2281 GGCTGTAGACAGACCCCAATTGATTTGGGATCTTTTCATATGTTATTGAAGTAAACAACTA 2340
Db      2396 GGCTGTAGACAGACCAATTGATTTGGGATCTTTTCATATGTTATTGAAGTAAACAACTA 2455
Qy      2341 AGAAGATAGCATGTTCTGTATTACAG 2366
Db      2456 AGAAGATAGCATGTTCTGTATTACAG 2481

RESULT 2
LOCUS   AK032528
DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length
enriched library, clone:6430582D06 product:McKusick-Kaufman
syndrome protein, full insert sequence.
ACCESSION AK032528.1 GI:26328342
VERSION AK032528
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitesunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-Format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
PUBMED 11076861
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
PUBMED 12056661
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,

```

```

Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/
FEATURES
Location/Qualifiers
1..2658
/organism="Mus musculus"
/mol_type="mrna"
/strain="C57BL/6J"
/db_xref="FANTOM DB:6430582D06"
/db_xref="taxon:10090"
/clone="6430582D06"
/sex="male"
/tissue_type="olfactory brain"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
759..2471
/note="unnamed protein product; McKusick-Kaufman syndrome
protein [MG] [MGI:1891836, GB|NM_021527, evidence: BLASTN,
99%, match=2657]
putative"
/codon_start=1
/protein_id="BAC27912.1"
/db_xref="GI:26328343"
/translation="MSRLBAKPSLCKTEPLTSEKVRSTLSVLKGVASCYGSGRLK
QHNGGCVTTTSSALLRNLSVTHPVLKILTSSVQNVHSCFSGGLFTALCCNL
INIRQLDTPATIKNLKYLISLCTSLYSKACSCRIIPVDFRSTFTLSLVSILTS
KPCAMLTRKETDHIGALIKLAFLLTIPSTEEWVIGKSIIVPLKQRTVDSITVPLG
LFEAGEVOLRLPTQKASGLVALFCTSLSGDFSNAGGVVVAHVQVSLNVALEOL
NLNGLRLVTHVDLVLCKVIHPSLKQFESRHHVMAIDRVGVTLMESLSKVTCATPIG
SLNPVITTYGSGVKDVCARFSGKHFHLLPNEAVTCTLLCSRNDTANBEELKLTQCT
AMHVQLTITKEPWLILGGCTTHLAAVYVRHVEAEAIVRDDGCTQAKLHVAAEAF
CSALESVAGSLSHDGGEILIDTKYHLNWSQADSASVGNWSDTSLRCGGLYNSQBEL
SWSVLRSYTHPPAPQTCLPQAALGSASNLTVDCFTAKLSGLQVAVETANLILDSLYVI
EDKN"
2639..2644
/polyA_signal
/note="putative"
2658
/polyA_site
/note="putative"
ORIGIN
Query Match 62.5%; Score 1544.2; DB 3; Length 2658;
Best Local Similarity 80.3%; Pred. No. 2.4e+304;
Matches 1889; Conservative 0; Mismatches 448; Indels 16; Gaps 6;
Qy 1 TATGAGCCTTCGGAACCTTGTGGAGAGACTACAAAGTTTGGTGTGTTATGCTCCCTTAGT 60
Db 147 TATGAGCTTTCAGATCTCTGGAGAGAGTACAAAGTTCTGATGTTATGATGATCATTAAT 206
Qy 61 TGGGCTCATACATTTGGGGTGGGTGATACAAATCAAAAGACGCCCTGTTTCCAAATACCTAA 120
Db 207 TGGGTTTCATACATTTGGGGTGGGCACAGAATCAAAACACAGCCCTGTCTTCCAAGTTCCT-- 264
Qy 121 AAACGACGACATCTCTGACAGACAGATAGTCTGGGACTTTCAAATCTTCAGAGAGCCCAAT 180
Db 265 -AAGGACGATACCACTGAGCTGGATAGTCTGGGACTTTCGAAGTCCACACGAGAGCAAC 323
Qy 181 CCAGGGGAAGTAGTAGCGGCTTGCAATCTTCAGGTAAAGAGCAGCTTGAATCTGAGCTTC 240

```


RESULT 3
AK032554
LOCUS
DEFINITION
AK032554 Mus musculus adult male olfactory brain cDNA, RIKEN full-length
enriched library, clone:6430593117 product:McKusick-Kaufman
syndrome protein, full insert sequence.
ACCESSION
VERSION AK032554.1 GI:26328364
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2658)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp,
URL: http://genome.gsc.riken.jp/; Tel: 81-45-503-9222,

Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/.
FEATURES
source
1. .2658
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:6430593117"
/db_xref="taxon:10090"
/clone="6430593117"
/sex="male"
/tissue_type="olfactory brain"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
759. .2471
/note="unnamed protein product; McKusick-Kaufman syndrome
protein (MGI:1891836, GB|NM_021527, evidence: BLASTN,
99%, match=2657)
putative"
/codon_start=1
/protein_id="BAC27921.1"
/db_xref="GI:26328365"
/translation="MSRLKAPKSLCKTPELTSEKVRSTLSVLKGVASCYGSGRLK
QUHNGGCVYTSOSSALLNRLSVTHPVKILTSVQNHVSCFSGGLFTALCCNL
INIRQLDTPATKLNKYLSSCTSYLKSACSCRI PVDPRSTHTFSLVHSILTS
KPCMLTKRETGHGALILKAFLLIPETERMVLGKSIILPLKQRTVDSVLGRL
LIEABEQLRLFTQKAGLRVALFCTSLSGDFSNAGGVVAVHVTQVSLNAVLSQL
LNLGRLVTHVDLVLCQKVIHPSKQFFSERHVAIDRVGTVLMESSKVTGATPIG
SLNPVSTYTGKQVCSARFGSKHFFHLLPNEATVCTLLGSRNDTAWELKLTQAT
AMHVLQITKEPWLGGCTTHLAAVVRHVHEAEI VRDDGCTQAKLHVAAAF
CSALRSVAGSLHDGGEILIDTKYGLHWSQADSASVGNWSDTLSCGGLYNSBEL
SNLSRLSYTHFPQTCLPQALGGSANLTVDFCTAKLSGLQAVETANLILDSYVI
EDKN"
polyA_signal 2639. .2644
/note="putative"
polyA_site 2658
/note="putative"
ORIGIN
Query Match 62.5%; Score 1544.2; DB 3; Length 2658;
Best Local Similarity 80.3%; Pred. No. 2.4e-304;
Matches 1889; Conservative 0; Mismatches 448; Indels 16; Gaps 6;
QY 1 TATGAGCCTTCGGAACCTTGTGGAGAGACTACAAAGTTTTGGTTGTTATGTCCTTAGT 60
DB 147 TATGAGCCTTCGGAATCTCTGGAGAGAGTACAAAGTTCTGATGGTTATGTTACCATTAAT 206
QY 61 TGGGGCTCATACATTTGGGGTGTGACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 120
DB 207 TGGGTTTCATACATTTGGGGTGTGACAGAAATCAAAAGCAGCCCTGTTTCAAGTTCT-- 264
QY 121 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTTCAAAATCTTCAGAAGAGCCAAAT 180
DB 265 -AAGGACGATACCACTGAGCTGGATAGTCTGGGACTTTCAGAGTCCACAGAGAGCCAAAC 323
QY 181 CAGGGGAGTAGCAGGCTTGCAATCTTCAGTAAAGAGCAGCTTTGAAATCTGAGCTTC 240
DB 324 CTAGGAGAGTAGTCCATCTTGCAAGTTTTCAG--AAAGAGCAACTTTGAAATCTGAACTTG 381
QY 241 ATATCGAAGAAGAGATGAAAATACCAGTTGGATTAGAAAGAACTGGCTTCTTGTAGCT 300
DB 382 ATATTAAGAAGAAGAAATGAAAACCCAGTTGGATTAGAAAGAACTGGCTTCTTGTGGCT 441
QY 301 GGGATATCTTTTCATAGGTGTCATCTTGGAAATATCTTTTTCGAGAGGTCTCGCAAGCAG 360
DB 442 GGGGTGACTTTTCATAGGCGTCCATCTTGGAAATATCTTTTATACAGAGAGTTCGAAAGAG 501

QY	361	TCTGTAAATTTTCAGTCTCAAGCAAA	CAAAAGAGATTTGAAGAGTGAAGTAAAT	420
Db	502	TCTGTAGGCTCTGAGCCAGAGCCAGCA	CAAAAGATTTGAAGATTTGAAGCAAAAT	561
QY	421	ATTGTGAATTTACTAAATTTGTTCATTA	ATATCTATGCTGATTAGCTTCATAA	480
Db	562	ACTTTGGATTATTA--TGTCATTAATTA	NAICTATGCTGGTTATTGTCTGAAC	619
QY	481	ACTTTTGGATTTTATAGCCACAATGCTG	CATATTCATATCTTAAATTTCTTAA	540
Db	620	GATTTTGTAGCTGGT-----AATGCTG	CATGTGCAGAC--TCATACCTGGAG	671
QY	541	TTTAATGTTAAACGTGATAATGCAATA	ATAAGAAAATGTGGTTTCAAAAATA	600
Db	672	TTTAATGTTTCAACACATGACAGTGC	AAATAAATGAAGCCATATGAACAA	731
QY	601	GTCTTCACCTAGTTTACCACCTGAAGT	AAAGATGTCTCGTTTGGAGCTTA	660
Db	732	GCCTTCACAAGTTA--CACCTGAAGC	AGGATGTCTCGTTTAGAAGCTTA	790
QY	661	GTGTAAAGTGAAACCACTGACAACTG	AGAGAGTCAGGACCACACTTTTCTG	720
Db	791	GTGTAAAACTGAAACCATTTGACAA	GTGAGAAAGTCAGTCCACACTTTT	850
QY	721	AATTGTAACATCATGCTATGGCCCTC	AGTAGGCTGAAGCAGCTGCACAA	780
Db	851	AGTCATAGCCCTCATGTATGGCCCTT	CAGGGAGCTTGAAGCAGCTACA	910
QY	781	AGGTTACGTGTGACAACTTCACAGT	CTCAGCTCTGCTCAGTCACCTTT	840
Db	911	AGGTTGTGTACAACTTCACAGTCTC	CAGCCCTGCTTCGAAACCTTT	970
QY	841	TCCCATTTTAAAGATCCTGACAGCCT	CCATPACAGAAATCATGTGTCA	900
Db	971	TCCCGTATTAAGATCCTTAAGCTCAT	CCGTGCGAATCAGTGTCTCTC	1030
QY	901	TGSCCTTATTCAGAGCTATTTCTT	GTGCGAACCTGATGAAATGTT	960
Db	1031	CGGCTTATTCACAGCCATTTCTGTG	CTGCAACTTTGATTGAAATTA	1090
QY	961	GACACCCACACATGCTATAGATTAA	ATAATAAATCTTTTGGCTTT	1020
Db	1091	GACACCCGACATGCTATTAATTA	ATAATAATACCTTGTAGTCT	1150
QY	1021	CAAGTCTGATPACTGTGGTGTGCA	ATCCAGTGGACTTTAGPAGTCT	1080
Db	1151	CAAGTCTGAAGCCCTGTAGTTGTG	CAATCCAGTTGACTTCAGATCT	1210
QY	1081	TTGTTTGGTGGCTAGTATATTA	CAAGTAAACCTGCTGTATGCT	1140
Db	1211	CAGCTTGGTACACAGTATCTTAA	CAAGCAACAGCCTGTATGCT	1270
QY	1141	AGAGCATGTCAGTGTCTTGATCTG	AGAGCCTTTTGTGCTTACAA	1200
Db	1271	AGATCATAGTGTCTTTGATTTGA	AGCTTTTACTTACAAATTC	1330
QY	1201	AGGCCACATCATTTTAGGAAAG	AGTTTAAATGTTAAAGGTC	1260
Db	1331	AGAACGAATGGTTTTAGGGAAG	AGTATAATTGTTCCCTTAA	1390
QY	1261	TTCCACTGTATTAACCTGGGAT	ACTCATTTGAATGTGAGAGTT	1320
Db	1391	TTCTACTGTATTAACCTGGACT	ACTCATTTGAAGCATCAGA	1450
QY	1321	ACCTATCAAAAATCAACTGCC	CTCAAGGTGGCACTCTTTT	1380
Db	1451	ACCCACTCAAAAGCGCAAGTG	GCCTTAGGGTGGCACTCTT	1510
QY	1381	CACCTTCGACATCGGAGAGGA	ACTGTGGTGTGAGTTATGGG	1440
Db	1511	TTTTCCTAATGCTGGAGAGG	AGTGTGGTGGCCCATTTAC	1570
QY	1441	AGCTCTGGACCAGCTGCTTAA	CTAGGAGGCAGCTAATCAGT	1500
Db	1571	AGTTTTAGAGCAGTTGCTTAA	TCTAGGAAGGCGACTAGT	1630
QY	1501	CCTGTGCCAAAAAGTTATACAT	CCCATCTTTTGAAGCAGTTT	1560
Db	1631	CCTGTGCCAAAAAGTTAATCC	CCCATCTTTTGAACAGTCTT	1690
QY	1561	TGCCATAGACAGAAATGGAGT	GACTCTGTGATGAAACCCCTG	1620
Db	1691	GGCCCATTCACACAGATTTGG	GGTGACTCTGATGGAATCTT	1750
QY	1621	GCCTATTGGATCCCTAGGCT	CAATATGCTCCTAATAGTTAT	1680
Db	1751	GCCTATTGGTCTCTAAACCC	CAATAGTTTCTACTATTAT	1810
QY	1681	CACTGCAAAATTTGGCTCC	CAATTTTTTTCATCTTAT	1740
Db	1811	CTCTGCAGATTTGGCTCC	AAACATTTTTTTCATCTT	1870
QY	1741	CAGCTTGTCTCTGCAACAG	AAATGACATCTGCCCTGG	1800
Db	1871	CACCTTGTCTCTGCGAG	CAGAAATGACACCGCTGG	1930
QY	1801	GAGGCACTGCATGTCTCTG	CACTTAACTCAAGAAC	1860
Db	1931	AACAGCAATGCGACTTTG	CACTTAACTCAAGAAC	1990
QY	1861	CTGTACTGAAATCTCAT	TTTGGCTGCATATATCAG	1920
Db	1991	CTGTACAGBAAACACACT	TTGGCTGCATATCTCAG	2050
QY	1921	CATTCTCAAAGATGATGA	ATGACTCAAAACAGAACTT	1980
Db	2051	TATTGTACAGATGATGG	GTGACTCAGGCAAAAGCT	2110
QY	1981	CAGTCCCTAGAAATCTG	TTGTTGGCTCTTTAGAA	2040
Db	2111	CAGTGTCTGGAGTCCG	TTGCTCTTTTGGAAAT	2170
QY	2041	CATGAAGTATGACACCT	TTTGGTCTCAGTTCAGG	2100
Db	2171	CACGAAGTATGACACCT	TTTGGTCTCTGCAAC	2230
QY	2101	AGATTTGCTTTCAGAGT	GTGGCTGTGGAATTA	2160
Db	2231	AGATACGCTGTCA	CGGTGTGGTGTGTTTGA	2290
QY	2161	TTTCTTAAAGAAGC	ACACGTCGTCCTATTTG	2220
Db	2291	TGCTTTAAGAGTACTT	ATCATCTTTTGCA	2350
QY	2221	GGGCTCAGCCAGCA	AACTTGACCTTTTGACT	2280
Db	2351	GGGCTCAGCCACT	TAACCTGACTGTGGA	2410
QY	2281	GGCTGTAGAGAC	CAGCAATTTGATTTGG	2340
Db	2411	GGCTGTAGAGAC	CAGCAATTTGATTTAG	2470
QY	2341	AGAGAAATAGCATG	2353	
Db	2471	ACATAAGAGAA	TG 2483	
RESULT 4				
AK004987				
LOCUS				
DEFINITION				
AK004987				
ACCESSION				
VERSION				
AK004987.1				
GI:12836588				
Mus musculus adult male liver cdna, RIKEN full-length enriched				
library, clone:1300013E18 product:McKusick-Kaufman syndrome				
protein, full insert sequence.				
AK004987				
2745 bp mRNA linear HTC 03-APR-2004				

KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	Carninci, P., and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M., Sumi, N., Iehii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuyama, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	
AUTHORS	The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
REFERENCE	
AUTHORS	Adachi, J., Aizawa, K., Carninci, P., Akimura, S., Arai, A., Aono, H., Arakawa, T., Bono, H., Kanihara, S., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirokawa, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	Please visit our web site (http://genome.gsc.riken.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer 15' GAGAGAGAGCGCCGACTCGATGTTTTTTTTTTTTTTT 3', cDNA was prepared by using trehalose thermo-activated reverse transcriptase

Db	557	ACTTTGGATTATTA--TGTCAATAAATAATCTATGTCGTGTTATTTCTCTGAACCACTT	614
Qy	481	ACTTTTTTGATTTTATAGCCCAATGCTGCATATTTATCATCTTTAATTCCTTAAGAATAATT	540
Db	615	GATTTTGTAGCTGTT-----AATGCTGCATGTGCAGAC---TCATACCTGGAGACCAGTC	666
Qy	541	TTTTAATGTTAAAACGTCATGATTAATGCAATAAATAGAAAAATGTGGTTTACAAAAATAAACG	600
Db	667	TTTTAATGTTCAAACACATGACAGTGCATAAATGNAAGCCATATGAACAACCTAAGACT	726
Qy	601	GTCCTTCACTAGTTTACCACTGAAGTAAGATGTCCTCGTTTGGAGCTTAAGAAGCCTCATTT	660
Db	727	GCCTTCCAAGTTA-CACCTGAAGCAGGATGTCCTCGGTTAGAAGCTAAGAAGCCATCGTT	785
Qy	661	GTGTAAAGAGTCAACCACTGACCAACTGAGAGAGTCAGGACCACACACTTTTCTCTTTGAAAAG	720
Db	786	GTGTAAACCTGAACCATTTGACAAGTGAGAAGTCAGGTCCACACTTTCTCTGTGAAAAGG	845
Qy	721	AATTTGTAACATCATGCTATGCCCCCTCAGGTAGGCTGAAGCAGCTGCACAATGGCTTTGG	780
Db	846	AGTCATAGCCTCATGCTATGGCCCTTCAGGGAGGCTGAGCAGCTACACATGGATTGGG	905
Qy	781	AGGTTACGTGTGTHACAACCTCACAGTCTCTCAGCTCTGCTCAGTCACTTTTGGTTCACACA	840
Db	906	AGGTTGTGTGTACACAACCTCACAGTCTCTCAGCCCTGCTTCGAAACCTTTTCAGTCAACCA	965
Qy	841	TCCCATTTTAAAGATCCTGACAGCCTCCATACAGAATCATGTGTCAAGCTTCAGTGATTCG	900
Db	966	TCCCGTATTAAAGATCTTAACGTCATCCGTGCAGAATCAGCTGTCTGCTTCAGTGACTG	1025
Qy	901	TGGCTTATTACAGCTATTCTTTGCTGCAACCTGATTGAAAATGTTTCAGAGATTTAGGCTT	960
Db	1026	CGGCTTATTACAGCCATTCTGTGCTGCACTTGATTTGAAAATATTCAAAGACTAGATTT	1085
Qy	961	GACACCCACCACTGCTCATTTAGATTAATAAACAFCPTTTTGAGTCTTTTGCATCAGTTATCT	1020
Db	1086	GACACCCGCACTGCTATTAAATTAATAATACCTCTTGAGTCTCTGTACCAGTTATCT	1145
Qy	1021	CAAGCTGTATACCTGTGGTTGTGCAATCCAGTGGACTTTTAGTAGTACTCAGATCCTCCT	1080
Db	1146	CAAGCTGGAAGCTGTAGTTGTGCAATCCAGATTTGACTTCAGAAGTACACATACCTTCCT	1205
Qy	1081	TTGTTTGGTGGTAGTATATTAAACAAGTAAACCTGCCCTGTATGCTACCCAGAAGGAAC	1140
Db	1206	CAGCTTGGTACACAGTATCTTAAACAGCAACACAGCTGTATGCTCACAGAAAGGAAC	1265
Qy	1141	AGACATGTCTAGTCTTTGATPCCTGAGAGCCTTTTTCCTTACAATTCAGAAAATGCTGA	1200
Db	1266	AGATCACATAGTGTCTTGATTTTGAAGCTTTTTTACTTACAATTCAGAAAGTACAGA	1325
Qy	1201	AGGCCACATCATTTTAGGAAGAGTTTAAATGTACTTTAAAGGTCAAAGAGTTATAGA	1260
Db	1326	AGAACGAATGGTTTTTAGGGAAGATTAATTTGTTTCCCTTAAAGGGCCANAAGAGTTACAGA	1385
Qy	1261	TTCCACTGTATTACTCGGATFACTCATTTGAAATGTCAGAAGTTCAAATTAATAGGCTATT	1320
Db	1386	TTTCTACTGTATTACTGGACTTACTCATTTGAAGCATCAGAAGTTCANNTAAGGAGTTATT	1445
Qy	1321	ACCTATCAAAAATAACAATGCCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCCGAGA	1380
Db	1446	ACCCACTCAAA-----	1456
Qy	1381	CACCTTCGACACTGGAGNAGGAACCTGGTGGTCAGTTATGGGGTTTCTCTTGAAAATGC	1440
Db	1457	-----	1456
Qy	1441	AGCTTTGGACCAGCTGCTTAACCTAGGAAGGCAGCTAATCAGTGACCACGCTAGATCTTGT	1500
Db	1457	-----	1456
Qy	1501	CCTGTGCCAAAAAGTTATPACATCCATCTTTGAAGCAGTTTCTCAATATGCATCGTATTAT	1560

Db	1457	-----AGTTCTTCAGTGAGCGCCACGTTAT	1481
Qy	1561	TGCCATAGACAGAAATTGGAGTGACTCTGATGGAAACCCCTGACTAAAAATGACAGGAAACACA	1620
Db	1482	GGCCATCGACAGAGTTGGGGTGACTCTGATGGAAATCTCTAAGCMAAGTGACAGGAGCAAC	1541
Qy	1621	GCCTATTGGATCCCTAGGCTCAATATGTCCTAATAGTTATGGAAGTGTGAAGATGTGTG	1680
Db	1542	GCCTATTGGTTCTCTAAACCCAATAGTTTCTACTACTTATGGAAGTGTGAAGATGTGTG	1601
Qy	1681	CATGCAAAATTTGGCTCCCAACATTTTTTTTCATCTATTTCTTAATGAAGCACAATCTG	1740
Db	1602	CTCTGCAAGATTTGGCTCCAAAATTTTTTTTCATCTTTCTTCTTAATGAGGCCACTGTCTG	1661
Qy	1741	CAGCTTCCTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCACGCTGTCAC	1800
Db	1662	CACCTTGCTTCTCTGCAAGAAATGACACCGCTGGGAGAGCTGAGCTCACATGTCA	1721
Qy	1801	GACGGCACTGCATGTCTCTGAGTTAAACACTCAAGGAACCAATGGGCTTTGTTGGAGGTGG	1860
Db	1722	AACAGCAATGCAGTCTTTCAGTTAAACAAATCAAGGAACCGTGGGTTTATTTGGAGGTGG	1781
Qy	1861	CTGTACTGAAACATCATTTGGCTGCATATATCAGACACAAAGCTCACAAACCCAGNAAG	1920
Db	1782	CTGTACAGAAACACACTTGGCTGCATATGTGCAGACACAAAGTTTCATCACAGCGCAGAAGC	1841
Qy	1921	CATTCTCAAAGATGATGAATGTACTCAAACAGAACTTCAATTAATTTCTGCAAGCATTTTG	1980
Db	1842	TATTTGTCAGAGATGATGGGTGTACTCAGGCAAGCTGCATGTTGCTGCTGAGGATTTTG	1901
Qy	1981	CAGTGCCTTAGAATCTGTTGTTGGCTCTTTAGAACATGATGGAGGTGAAATTTCTCACTGA	2040
Db	1902	CAGTGCCTCTGGAGTCCGTTCTGGCTCTTTTGGAAACATGATGGTGGTGAATTCCTCATTTGA	1961
Qy	2041	CATGAAGTATGGACACCTTTGGTTCAGTTCAGGCAGATTTCTCCCTGTGTGTTCTTAACGGCC	2100
Db	1962	CACGAAGTATGGACACCTTTGGTCTGTCTCAAGCAGATTTCTGCGCTGTGTGGTAACTGGTC	2021
Qy	2101	AGATTTGCTTTTTCAGGTGCGCTGTGGATTATACAATPAGCCAGAGAACTCAACTGGTC	2160
Db	2022	AGATACGCTGTACGGTGTGGCTGTGGTTTGTACAACAGCCAGAGAGCTCAGCTGGTC	2081
Qy	2161	TTTTCTTAAGAACACACAGTCGTCCATTTGTGCCACAAAGCTGCTTCCACATGAAGCTGT	2220
Db	2082	TGCTTTAAGAAGTACTTTATCATCTCTTTTGACACACAAACCTGCTTCCACAGGCAGCTTT	2141
Qy	2221	GGGCTCAGCAGCAACCTGACCTTGGACTGTTTGAAGCTGCAAGCTTAGTGGCCCTACAGGT	2280
Db	2142	GGGCTCAGCAGTAACCTGACTGTGGACTGCTTTCACTGCCAAGCTGAGTGGCTTTACAGGT	2201
Qy	2281	GGCTGTAGACAGCCAAATTTGATTTGGGATCTTTTATATGTTATTGAAGATAAAAACTA	2340
Db	2202	GGCTGTAGACAGCCAAATTTGATTTTATGATCTTTTATGATCTATTGAAGATAAAAACTA	2261
Qy	2341	AGAGAATAGCATG 2353	
Db	2262	ACATAAGAGAAATG 2274	
RESULT 5			
BM473886			
LOCUS	BM473886	1105 bp	mRNA
DEFINITION	AGENCOURT 6484160 NIH_MGC_72 Homo sapiens	linear	EST 05-FEB-2002
VERSION	5, mRNA sequence.		
ACCESSION	BM473886		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	1 (bases 1 to 1105)		
	NIH-MGC http://mgc.nci.nih.gov/.		

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTP/DTF

cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

High quality sequence stop: 04
Location/Qualifiers

FEATURES

source

1. .1105
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5538267"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_72"
/note="Organ: skin; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 2 kb. Library constructed by Life Technologies."

ORIGIN

Query Match 35.68; Score 880.8; DB 4; Length 1105;
Best Local Similarity 95.08; Pred. No. 4.7e-169;
Matches 963; Conservative 0; Mismatches 44; Indels 7; Gaps 5;

QY 1 TATGAGCCTTCGGAACCTGTGGAGAGACTACAAAGTTTGTGTTATGTCCTCTTAGT 60
DB 98 TATGAGCCTTCGGAACCTGTGGAGAGACTACAAAGTTTGTGTTATGTCCTCTTAGT 157
QY 61 TGGGCTCATACATTTGGGGTGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 120
DB 158 TGGGCTCATACATTTGGGGTGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 217
QY 121 AAAGCAGCATTCCTGAGCAAGATAGTCGGACATTTCAAACTTCAGAAAGCCAAAT 180
DB 218 AAAGCAGCATTCCTGAGCAAGATAGTCGGACATTTCAAACTTCAGAAAGCCAAAT 277
QY 181 CCAGGGGAAGTAGCAGCCTTCGCAATCTTCAGGTAAAGACAGCTTTGAATCTGAGCTTC 240
DB 278 CCAGGGGAAGTAGCAGCCTTCGCAATCTTCAGGTAAAGACAGCTTTGAATCTGAGCTTC 337
QY 241 ATATCGAAAGAGAGATGAAATAATACAGTTGGATTAGAAAGAACTGGCTTCTTTGAGCT 300
DB 338 ATATCGAAAGAGAGATGAAATAATACAGTTGGATTAGAAAGAACTGGCTTCTTTGAGCT 397
QY 301 GGGATATCTTTTATAGTGTCCATCTTGGAAACATACCTTTTTCAGAGGCTCTGAAAGCAG 360
DB 398 GGGATATCTTTTATAGTGTCCATCTTGGAAACATACCTTTTTCAGAGGCTCTGAAAGCAG 457
QY 361 TCTGTAATAATTTTCAGTCTCAAGCAACAAAGAGATTTCAGAGGTGAAGTAAATAAAT 420
DB 458 TCTGTAATAATTTTCAGTCTCAAGCAACAAAGAGATTTCAGAGGTGAAGTAAATAAAT 517
QY 421 ATTTGGAATTTACTAATTTGTTCATTAATCATTTCTATGCTGATTAGCTTCATAAATGGA 480
DB 518 ATTTGGAATTTACTAATTTGTTCATTAATCATTTCTATGCTGATTAGCTTCATAAATGGA 577
QY 481 ACTTTTGTATTTATAGCCCAATAGCTGCAATTCATATCTTAATTTAACTTCTTAAAGAAAT 540
DB 578 ACTTTTGTATTTATAGCCCAATAGCTGCAATTCATATCTTAATTTAACTTCTTAAAGAAAT 637
QY 541 TTTTAATGTTAAACGTTGATATGCAATAATAGAAAATGCTGTTTACAAAATAAAGC 600
DB 638 TTTTAATGTTAAACGTTGATATGCAATAATAGAAAATGCTGTTTACAAAATAAAGC 697
QY 601 GTCTTCACTAGTTACCACTGAAAGTAGATGTCGTTTGGAAAGCTTAAGAGCCATCAT 660

DB 698 GTCTTCACTAGTTACCACTGAAAGTAGATGTCGTTTGGAAAGCTTAAGAGCCATCAT 757
QY 661 GTCTAGAGTGAACCACTGACAACTGAGAGAGTCAAGACACACACTTCTCTCTTGAAG 720
DB 758 GTCTAGAGTGAACCACTGACAACTGAGAGAGTCAAGACACACACTTCTCTCTTGAAG 817
QY 721 AATTGTAACATCATGCTATGCGCCCTCAGTAGAGCTGGAAGCAGCTGCACAAAT-GGCTTTG 779
DB 818 AATTGTAACATCATGCTATGCGCCCTCANGTANGCTGAAGCAGCTGCACAAATGGGCTTTG 877
QY 780 GAGTTACGTTGTACAACCTCAAGCTCAGCTCTGCTCAGCTCAGCTCTTGGTTCACAC 839
DB 878 GAGTTACGTTGTACAACCTCAAGCTCAGCTCTGCTCAGCTCAGCTCTTGGTTCACAC 937
QY 840 ATCCCAATTTTAAAGATCCTGACAGCCTCCATACAGAAATCATGTCTCAAGCTTCAGTGA 899
DB 938 CTCCCAATTTTAAAGATCCTGACGCTCCATAC-GAATCATGTGTCAAGCTTCAAGGAT 994
QY 900 GTGGCTTATTCAGAGCTATTTTGTGCAACCTGATTGAAATGTTTCAGAGATTAGGCT 959
DB 995 GGTGGCTTATTCAGAGCTATTTTGTGCAACCTGATTGAAATGTTTCAGAGATTAGGCT 1054
QY 960 TGACACCCCACTGTCATTAGATTAAATAAACAATCTTTTGAAGTCTTTTGCATCA 1013
DB 1055 TGACA-CCACCACCTGTCATTAAAT--AATAAATCTCTTTTGGAGTCTTGCATCA 1105

RESULT 6
BM558221
LOCUS BM558221 1019 bp mRNA linear EST 20-FEB-2002
DEFINITION AGENCOURT_6588340 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5477987
5', mRNA sequence.

ACCESSION BM558221
VERSION BM558221.1 GI:18800793
KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov

Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLM1996 row: 0 column: 12
High quality sequence stop: 726.

Location/Qualifiers

1. .1019
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5477987"

/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_98"

/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene), and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

FEATURES
source

ORIGIN

Query Match		35.2%;	Score 870.4;	DB 4;	Length 1019;
Best Local Similarity		97.0%;	Pred. No. 6.2e-167;		
Matches 908;		Conservative 0;	Mismatches 26;	Indels 2;	Gaps 2;
QY	844	CATTTTAAAGATCCTGACAGCCTCCATACAGAAATCATGTGTCAAGCTTCAGTGTGCG	903		
DB	1	CATTTTAAAGATCCTGACAGCCTCCATACAGAAATCATGTGTCAAGCTTCAGTGTGCG	60		
QY	904	CTTATTACAGCTATTCTTTTGCTGCAACCTGATGAAAAATGTTACAGATTAAGCTTGAC	963		
DB	61	CTTATTACAGCTATTCTTTGCTGCAACCTGATGAAAAATGTTACAGATTAAGCTTGAC	120		
QY	964	ACCCACACCTGCTATTAGATAAATAAATCAATCTTTTGAGTCTTTGATCAGTATCTCAA	1023		
DB	121	ACCCACACCTGCTATTAGATAAATAAATCAATCTTTTGAGTCTTTGATCAGTATCTCAA	180		
QY	1024	GTCTGATACCTGTGGTTGTGGAATCCAGTGGACTTTAGTAGTACTCAGATCCTCTTTG	1083		
DB	181	GTCTGAGACTGTGGTTGTGGAATCCAGTGGACTTTAGTAGTACTCAGATCCTCTTTG	240		
QY	1084	TTTGGTGGCTGATATTAAACAGTAAACCTGCTGTATGCTCACCGAAAGGAAACAGA	1143		
DB	241	TTTGGTGGCTGATATTAAACAGTAAACCTGCTGTATGCTCACCGAAAGGAAACAGA	300		
QY	1144	GCATGTCAGTCTTTCATCTGAGAGCCTTTTGTCTCAATTCAGAAATGCTGAAAG	1203		
DB	301	GCATGTCAGTCTTTCATCTGAGAGCCTTTTGTCTCAATTCAGAAATGCTGAAAG	360		
QY	1204	CCACATCATTTTAAAGAAAGTTTAAATGTGACCTTTTAAAGGTCAAGAGTTATAGATT	1263		
DB	361	CCACATCATTTTAAAGAAAGTTTAAATGTGACCTTTTAAAGGTCAAGAGTTATAGATT	420		
QY	1264	CACGTATTACCTGGGATCTCATTTGAAATGTCAGAAATTCATTAATGAGGCTATTACC	1323		
DB	421	CACGTATTACCTGGGATCTCATTTGAAATGTCAGAAATTCATTAATGAGGCTATTACC	480		
QY	1324	TATCAAAAATCACTGCCCTCAAGTGGCACTTTTGTACAACTTTATCCGGACAC	1383		
DB	481	TATCAAAAATCACTGCCCTCAAGTGGCACTTTTGTACAACTTTATCCGGACAC	540		
QY	1384	TTCTGACACTGGAGAGGAACCTGCTGTGTCAGTATGGGTTTCTCTTGAAATGCAAT	1443		
DB	541	TTCTGACACTGGAGAGGAACCTGCTGTGTCAGTATGGGTTTCTCTTGAAATGCAAT	600		
QY	1444	CTTGGACCAAGCTGCTTAACTAGGAAAGGACGCTTAATCAGTGACCAAGTATCTTCT	1503		
DB	601	CTTGGACCAAGCTGCTTAACTAGGAAAGGACGCTTAATCAGTGACCAAGTATCTTCT	660		
QY	1504	GTGCCAAAAGTTATACATCCATCTTTGAAGCAGTTTCTCAATATCATCTGATTTATGC	1563		
DB	661	GTGCCAAAAGTTATACATCCATCTTTGAAGCAGTTTCTCAATATCATCTGATTTATGC	720		
QY	1564	CATAGACAGAAATGGAGTGACTCTGATGGAACCCCTGACTTAAATGACAGGACACAGC	1623		
DB	721	CATAGACAGAAATGGAGTGACTCTGATGGAACCCCTGACTTAAATGACAGGACACAGC	780		
QY	1624	TATTGGATCCCTAGGCTCAATATGCTTAATAGTTATGGAAGTGTGAAAGTGTGTCAC	1683		
DB	781	TATTGGATCCCTAGGCTCAATATGCTTAATAGTTATGGAAGTGTGAAATGTTGTCAC	840		
QY	1684	TGCAAAATTTGGCTCCCAACATTTTTTTCAT-CTTATCTCTAAT-GAAGCAACATCTGC	1741		
DB	841	TGCAAAATTTGGCTCCCAACATTTTTTTCATCTTCTCTTATCTTAAATGGAAGCAACCTGC	900		
QY	1742	AGCTTGCTTCTCTGCAACAGAAATGACACTGCTGCTG 1777			
DB	901	AAGCTTGCTTCTCTGCAACAAATAATGACCCCG 936			

RESULT 7
BQ641810
LOCUS

DEFINITION	AGENCOURT_8290193 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:6305558				
ACCESSION	BQ641810				
VERSION	BQ641810.1				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 920)				
TITLE	NIH-MGC http://mgc.nci.nih.gov/.				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: c9apbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Ruben Laboratory DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLCM2528 row: a column: 15 High quality sequence stop: 625.				
FEATURES	Location/Qualifiers				
source	1..920				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:6305558"				
	/tissue_type="normal pigmented retinal epithelium"				
	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH_MGC_43"				
	/note="Organ: eye; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library. "				
ORIGIN					
Query Match	34.4%; Score 849.8; DB 5; Length 920;				
Best Local Similarity	97.2%; Pred. No. 9.8e-163;				
Matches 896;	Conservative 0; Mismatches 23; Indels 3; Gaps 3;				
QY	1249	AAGAGTTATAGATTCCACTGTATTACCTGGGATCTCAATTGAAATGTCAGAAATTC	1308		
DB	1	AAGAGTTATAGATTCCACTGTATTACCTGGGATCTCAATTGAAATGTCAGAAATTC	60		
QY	1309	AATGAGGCTATTACCTATCAAAAAATCAAACTGCCCTCAAGGTGGCACTCTTTTGTACAAC	1368		
DB	61	AATGAGGCTATTACCTATCAAAAAATCAAACTGCCCTCAAGGTGGCACTCTTTTGTACAAC	120		
QY	1369	TTTATCCGGAGACACTTCTGCACATGGAGAGCACTGTGTGTCAGTTATGGGTTTC	1428		
DB	121	TTTATCCGGAGACACTTCTGCACATGGAGAGCACTGTGTGTCAGTTATGGGTTTC	180		
QY	1429	TCTTGAATAATGCAGTCTTCTGGACAGCTGCTTAACTTAGGAAGCAGCTAATCAGTGACCA	1488		
DB	181	TCTTGAATAATGCAGTCTTCTGGACAGCTGCTTAACTTAGGAAGCAGCTAATCAGTGACCA	240		
QY	1489	CGTAGATCTTGTCTGTGCAAAAAAGTTATATCATCATCTTTGAAGCAGTTTCTCAATAT	1548		
DB	241	CGTAGATCTTGTCTGTGCAAAAAAGTTATATCATCATCTTTGAAGCAGTTTCTCAATAT	300		
QY	1549	GCATGCTATTATTGCCATAGACAGATTTGGAGTGACTCTGTATGGACCCCTGACTAAAAT	1608		
DB	301	GCATGCTATTATTGCCATAGACAGATTTGGAGTGACTCTGTATGGACCCCTGACTAAAAT	360		
QY	1609	GACAGGAAACACAGCCTATTGGATCCCTAGGCTCAATATATCTCTTAATAGTTATGGAGTGT	1668		

Db	957	CAWCCCATTTTAAKATCTWKAYAK-CWACAWACAATAATCAATKTAKCAAKCWTCAKTKAT	1015
Qy	899	TGTGGCTTATTCACAGCTATTCTTTGCTGCAACCTGATTGAAATGTTTCAGAGATTAGGC	958
Db	1016	TKTKTCTAAT-CAMAGCWATCTTTGCTKCAACATGATK-AWAATKTCAGAGATATGYT	1071
Qy	959	TTGACACCCACCACTGTCATTAGATTAAATA	989
Db	1072	TKCAMCWCACAKTCATARGWTAAATAAAAAAHM	1102
RESULT 9			
LOCUS	BQ675098	930 bp	mRNA
DEFINITION	AGENCOURT 8349137 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6275978		EST 15-JUL-2002
ACCESSION	BQ675098		5', mRNA sequence.
VERSION	BQ675098	1	GI:21785932
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 930)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov		
	Plate: LLCM2460 row: a column: 03		
	High quality sequence stop: 616.		
FEATURES	Location/Qualifiers		
source	1..930		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:6275978"		
	/tissue_type="epidermoid carcinoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/clone_lib="NIH_MGC_102"		
	/notes="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI; Site 2: EcoRI; cDNA made by oligo-dr priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."		
ORIGIN			
Query Match	34.1%;	Score 843.4;	DB 5; Length 930;
Best Local Similarity	97.8%;	Pred. No. 2e-161;	
Matches	886;	Conservative 0;	Mismatches 17; Indels 3; Gaps 3;
Qy	484	TTTTTGATTTTATAGCCACAAATGCTGCATATTCATATCTTAATTCCTCAAGAATAATTTTT	543
Db	1	TTTTTGATTTTATAGCCACAAATGCTGCATATTCATATCTTAATTCCTCAAGAATAATTTTT	60
Qy	544	AATGTTAAACCGTGATAATGCAATAAATAGAAAAATGTGGTTTACAAAATAAAAACGGTC	603
Db	61	AATGTTAAACCGTGATAATGCAATAAATAGAAAAATGTGGTTTACAAAATAAAAACGGTC	120
Qy	604	TTCACTAGTTTACCACTGGAAGTGAAGTGTCTGTTTGAAGCTAAGAGCCATCATTTGTG	663
Db	121	TTCACTAGTTTACCACTGGAAGTGAAGTGTCTGTTTGAAGCTAAGAGCCATCATTTGTG	180

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: L10M698 row: d column: 20
 High quality sequence stop: 615.
 Location/Qualifiers

FEATURES

1. .932
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6536612"
 /tissue_type="epidermoid carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_101"
 /note="Organ: lung; Vector: pOTB7; Site_1: EcoRI; Site_2:
 XhoI; cDNA made by oligo-dr priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."

ORIGIN

Query Match 32.3%; Score 797.8; DB 5; Length 932;
 Best Local Similarity 99.5%; Pred. No. 4e-152;
 Matches 82; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
 794 ACAACCTCACAGTCTCTGCTCAGTCACCTTTTGGTCACACATCCCATTTAAAG 853
 Db 1 ACAACCTCACAGTCTCTGCTCAGTCACCTTTTGGTCACACATCCCATTTAAAG 60
 854 ATCTGACACCTCCATACAGATCATGTCTCAAGCTTCAGTATGTCCTATTACACA 913
 Db 61 ATCTGACACCTCCATACAGATCATGTCTCAAGCTTCAGTATGTCCTATTACACA 120
 914 GCTATTCTTGTGCAACCTGATTGAAATGTTTCAGAGATTAGGCTTGCACCCACCACT 973
 Db 121 GCTATTCTTGTGCAACCTGATTGAAATGTTTCAGAGATTAGGCTTGCACCCACCACT 180
 974 GTCATTAGATTAAATAACATCTTTTGGAGTCTTTGTCATCAGTTATCTCAAGTCTGATACC 1033
 Db 181 GTCATTAGATTAAATAACATCTTTTGGAGTCTTTGTCATCAGTTATCTCAAGTCTGAGACC 240
 1034 TGTGGTGTGCAATCCAGTGGACTTTAGTAGTACTCAGATCTCTCTTTGTTGGTGCCT 1093
 Db 241 TGTGGTGTGCAATCCAGTGGACTTTAGTAGTACTCAGATCTCTCTTTGTTGGTGCCT 300
 1094 AGTATATTACAAAGTAAACCTGCTGTATGCTCACCAGAAAGCAACAGACGATGTCAGT 1153
 Db 301 AGTATATTACAAAGTAAACCTGCTGTATGCTCACCAGAAAGCAACAGACGATGTCAGT 360
 1154 GCTTTGATCTGAGAGCTTTTGTCTTACAAATCCAGAAATGCTGAAGGCCACATCAATT 1213
 Db 361 GCTTTGATCTGAGAGCTTTTGTCTTACAAATCCAGAAATGCTGAAGGCCACATCAATT 420
 1214 TTAGGAAGAGTTTAAATGTACCTTTTAAAGGTCAAGAGTTATAGATTCACCTGATTA 1273
 Db 421 TTAGGAAGAGTTTAAATGTACCTTTTAAAGGTCAAGAGTTATAGATTCACCTGATTA 480
 1274 CCTGGGATCTCAATGAAATGTCAGAGTTCAATTAATGAGGCTATTACTATCAAAAAA 1333
 Db 481 CCTGGGATCTCAATGAAATGTCAGAGTTCAATTAATGAGGCTATTACTATCAAAAAA 540
 1334 TCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTTATCCGGAGACACTTCTGACACT 1393
 Db 541 TCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTTATCCGGAGACACTTCTGACACT 600
 1394 GGAGAGAGGAACCTGTGGTGGTCAATTATGGGGTTTCTTTGAAATGCAAGTCTTGGACCAAG 1453
 Db 601 GGAGAGAGGAACCTGTGGTGGTCAATTATGGGGTTTCTTTGAAATGCAAGTCTTGGACCAAG 660

1454 CTGCTTAACCTAGGAGGCACTAATCAGTGACACGACGTAGATCTTGTCTGTGCAAAAA 1513
 Db 661 CTGCTTAACCTAGGAGGCACTAATCAGTGACACGACGTAGATCTTGTCTGTGCAAAAA 720
 1514 GTTATACAT-CCATCTTTGAAGCAGTTTCTCAATATGTCATCGTATTATTATGCCATAGACAG 1572
 Db 721 GTTATACATCCCATCTTTGAAGCAGTTTCTCCATATGTCATCGTATTATTATGCCATAGACAG 780
 1573 -AATTGGAGTGACTCTGATGGAACCCCTGACTAAATGACAGGAA 1616
 Db 781 AAATTGGAGTGACTCTGATGGAACCCCTGACTAAATGACAGGAA 825

RESULT 11

BG742334 810 bp mRNA linear EST 15-MAY-2001
 602631495F1 NCI_CGAP_skn3 Homo sapiens cDNA clone IMAGE:4776791 5',
 mRNA sequence.

DEFINITION

BG742334
 BG742334.1 GI:14052987

ACCESSION

EST.
 SOURCE Homo sapiens (human)

KEYWORDS

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 810)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)

AUTHORS

Unpublished (1999)

TITLE

Contact: Robert Strausberg, Ph.D.

JOURNAL

Email: cgapbs-r@mail.nih.gov

COMMENT

Tissue Procurement: James Cleaver, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 Sequencing by: The I.M.A.G.E. Consortium (LLNL) DNA
 Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L10M10629 row: n column: 24
 High quality sequence stop: 807.

FEATURES

Location/Qualifiers
 1. .810
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4776791"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NCI CGAP_Skn3"
 /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
 Average insert size 1.5kb. Library constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 31.8%; Score 785.4; DB 4; Length 810;
 Best Local Similarity 99.6%; Pred. No. 1.3e-149;
 Matches 808; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
 1333 ATCAACTGCCCTCAA-GGTGGCACTCTTTTGTACAACTTTATCCGGAGACACTTCTGACA 1391
 Db 1 ATCAACTGCCCTCAAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGACACTTCTGACA 60
 1392 CTGGAGAAGAACTGTGGTGGTCAAGTTATGGGGTTTCTCTTGAATGCAAGTCTTGGACC 1451
 Db 61 CTGGAGAAGAACTGTGGTGGTCAAGTTATGGGGTTTCTCTTGAATGCAAGTCTTGGACC 120
 1452 AGCTGCTTAACTAGGAAGGAGCTAATCAGTGACACCTAGATCTTGTCTGTGCAAAA 1511
 Db 121 AGCTGCTTAACTAGGAAGGAGCTAATCAGTGACACCTAGATCTTGTCTGTGCAAAA 180
 1512 AAGTTATACATCCATCTTTGAAAGCAGTTTCTCAATATGATCGTATTATTATGCCATAGACA 1571

Db	181	AAGTTATACATCCATCTTTGAGCAGCTTTCTCAATATGATCGTATTATTATGGCCATAGACA	240
QY	1572	GAATTCGAGTGAATCTGATGGAACCCCTGACTAAATATGACAGAAACACAGCCCTATTGGAT	1631
Db	241	GAATTCGAGTGAATCTGATGGAACCCCTGACTAAATATGACAGAAACACAGCCCTATTGGAT	300
QY	1632	CCCTAGGCTCAATATGCTCTATAGTATGGAAGTGTGAAGATGTGCACATGCAAAAT	1691
Db	301	CCCTAGGCTCAATATGCTCTATAGTATGGAAGTGTGAAGATGTGTGCACATGCAAAAT	360
QY	1692	TTGGCTCCCAACATTTTTTTTCATCTTATTCCTAATGAAGCAACAATCTGCAGCTTCTCTC	1751
Db	361	TTGGCTCCCAACATTTTTTTTCATCTTATTCCTAATGAAGCAACAATCTGCAGCTTCTCTC	420
QY	1752	TTTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCAGCTGTGACAGCGCACTGC	1811
Db	421	TTTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCAGCTGTGACAGCGCACTGC	480
QY	1812	ATGCTCTGCGAGTTAACTCAAGGAACCATGGGCTTTGTTGGAGGTGGCTGTACTGAA	1871
Db	481	ATGCTCTGCGAGTTAACTCAAGGAACCATGGGCTTTGTTGGAGGTGGCTGTACTGAA	540
QY	1872	CTCATTTGGCTGCATATATCAGACACAGACTCACAAACGCCAGAAAGCATTTCTCAAAG	1931
Db	541	CTCATTTGGCTGCATATATCAGACACAGACTCACAAACGCCAGAAAGCATTTCTCAAAG	600
QY	1932	ATGATGAATGTACTCAACAGAACTTCAATTAATTTGCTGAAGCATTTTGGAGTGCCTAG	1991
Db	601	ATGATGAATGTACTCAACAGAACTTCAATTAATTTGCTGAAGCATTTTGGAGTGCCTAG	660
QY	1992	AATCTGTGTGTGCTCTTTAGAACATGATGGAGGTGAAATTTCTCACTGACATGAACTATG	2051
Db	661	AATCTGTGTGTGCTCTTTAGAACATGATGGAGGTGAAATTTCTCACTGACATGAACTATG	720
QY	2052	GACACCTTTGGTCAGTTCAGGCAGATTTCTCCCTGTGTGTGCTAACTGGCCAGATTTGCTTT	2111
Db	721	GACACCTTTGGTCAGTTCAGGCAGATTTCTCCCTGTGTGTGCTAACTGGCCAGATTTGCTTT	780
QY	2112	CACAGTGTGCTGTGATTAATACANTAGCCA	2142
Db	781	CACAGTGTGG-TGTGGATTATACAATAGCCA	810
RESULT 12			
BM913546			
LOCUS			
DEFINITION	BM913546 1044 bp mRNA linear EST 12-MAR-2002		
ACCESSION	AGENCOURT_6610292 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5475679		
VERSION	5', mRNA sequence.		
KEYWORDS	BM913546		
SOURCE	EST. BM913546.1 GI:19363925		
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
JOURNAL	1 (bases 1 to 1044)		
COMMENT	NIH-MGC http://mgc.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: DCTD/DTF cDNA Library Preparation: Rubin Laboratory DNA Sequencing by: The I.M.A.G.E. Consortium (LNL) Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LLCMI990 row: o column: 08 High quality sequence stop: 679. Location/Qualifiers 1. .1044 /organism="Homo sapiens"		
FEATURES			
source			

ORIGIN

Query Match	31.6%;	Score 781;	DB 5;	Length 1044;
Best Local Similarity	98.5%;	Pred. No. 1.1e-148;		
Matches 798;	Conservative 0;	Mismatches 11;	Indels 1;	Gaps 1;
QY	1612	AGGAACACAGCCCTATTGGATCCCTAGGCTCAATATGCTCTAATAGTTATGGAAGTGTGA	1671	
Db	1	AGGAACACACAGCCCTATTGGATCCCTAGGCTCAATATGCTCTAATAGTTATGGAAGTGTGA	60	
QY	1672	AGATGTGTGCACCTGCAAAAATTTGGCTCCCAACATTTTTTTTCATCTTATTCCTTAATGAAGC	1731	
Db	61	AGATGTGTGCACCTGCAAAAATTTGGCTCCCAACATTTTTTTTCATCTTATTCCTTAATGAAGC	120	
QY	1732	AACAATCTGCAGCTTGTCTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCT	1791	
Db	121	AACAATCTGCAGCTTGTCTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCT	180	
QY	1792	CAGGTGTGAGAGGCACTGCATGCTCTGCAGTTTAACTCAAGGAACCAATGGCTTTGTT	1851	
Db	181	CAGGTGTGAGAGGCACTGCATGCTCTGCAGTTTAACTCAAGGAACCAATGGCTTTGTT	240	
QY	1852	GGAGGTGGCTGTACTGAAACTCATTTGGCTGCATATATCAGACACAGACTCACAACGA	1911	
Db	241	GGAGGTGGCTGTACTGAAACTCATTTGGCTGCATATATCAGACACAGACTCACAACGA	300	
QY	1912	CCCAGAAAGCATTTCTCAAGAGATGATGAATGTACTCAACAGAACTTCAATTAATTCCTGA	1971	
Db	301	CCCAGAAAGCATTTCTCAAGAGATGATGAATGTACTCAACAGAACTTCAATTAATTCCTGA	360	
QY	1972	AGCATTTTGCAGTGCCTTAGAATCTGTTGGCTCTTTAGAACATGATGGAGTGAAT	2031	
Db	361	AGCATTTTGCAGTGCCTTAGAATCTGTTGGCTCTTTAGAACATGATGGAGTGAAT	420	
QY	2032	TCTCACTGACATGAAGTATGGACACCTTTGGTTCAGTTTCAAGGAGATTTCTCCCTGTGTTGC	2091	
Db	421	TCTCACTGACATGAAGTATGGACACCTTTGGTTCAGTTTCAAGGAGATTTCTCCCTGTGTTGC	480	
QY	2092	TAACTGGCCAGATTTTCTTTCACAGTGTGGCTGTGGATTTATACAATAGCAGGAAGAACT	2151	
Db	481	TAACTGGCCAGATTTTCTTTCACAGTGTGGCTGTGGATTTATACAATAGCAGGAAGAACT	540	
QY	2152	CAACTGGCTTTTCTTAAGAGACACAGCTCGTCCATTTGTGCCCAAGCTGCTTCCACA	2211	
Db	541	CAACTGGCTTTTCTTAAGAGACACAGCTCGTCCATTTGTGCCCAAGCTGCTTCCACA	600	
QY	2212	TGAAGCTGTGGCTCAGCCAGCAACCTTGACTTGGACTGTTTGGCTGCAAAAGCTTAGTGG	2271	
Db	601	TGAAGCTGTGGCTCAGCCAGCAACCTTGACTTGGACTGTTTGGCTGCAAAAGCTTAGTGG	660	
QY	2272	CCTACAGGTGGCTGTAGACAGCCAAATTTGATTTTGGGATCTTTTCATATGTTATTGAAGA	2331	
Db	661	CCTACAGGTGGCTGTAGACAGCCAAATTTGATTTTGGGATCTTTTCATATGTTATTGAAGA	720	
QY	2332	TAAAACTAAGAGATAGCATGTTTGGTATATACAGAGAACAA-ATAAACTAGTCTGTTG	2390	
Db	721	TAAAACTAAGAGATAGCATGTTTGGTATATACAGAGAACAA-ATAAACTAGTCTGTTG	780	
QY	2391	GCAATTTGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG	2420	

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5475679"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_41"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GCCACGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

/note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN		
Query Match 30.4%; Score 752.2; DB 5; Length 942;		
Best Local Similarity 92.1%; Pred. No. 8e-143;		
Matches 856; Conservative 0; Mismatches 64; Indels 9; Gaps 6;		
QY	940	AAATGTTTCAGAGATTAGGCTTTGACACCCACCACTGTCATTAGATTAAATAAATCAATCTTTT 999
DB	922	AAATGTTCAAAATAGGCTTGACCAACCACTGTCAT-AGATTAAATAAATCAATCTTT- 865
QY	1000	GAGTCTTTGCATCAGTTATCTCAAGTCTGATCACTGTGGTGTGCGAATCCAGTGACGTT 1059
DB	864	--GAGTCTTGATCAGTATTTCAAGTCTGAGACCTGGGTT-TCGAATCCAGTGACCTT 808
QY	1060	TAGTAGTACTCAGATCTCTCTTTGTTGGTGGCTAGTATATTA-ACAAGTAAACCTGCCT 1118
DB	807	TAGTAGACTCAAGATCTCTCTTTGTTGGTGGCTAGTATATTAACAAGTAAACCTGCCT 748
QY	1119	GTATGCTCACCAAGAAAGAACAGACATGTCAGTCTTTGATCCTGAGAGCCTTTTTCG 1178
DB	747	GTATGCTCACCAAGAAAGAACAGACATGTCAGTCTTTGATCCTGAGAG-CTTTTTCG 689
QY	1179	TTACAATTCAGAAAATGCTGAAGGCCACATCAATTTTAGAAAAGATTTAATTTGTAACCTT 1238
DB	688	TTACAATTCAGAAAATGCTGAAGGCCACATCAATTTTAGGANAGATTTA--TTGTACTT 631
QY	1239	TAAAGGTCAAGAGTTATAGATTCACCTGTATTAATCTGGGATACCTATTGAATGTCAG 1298
DB	630	TAAAGGTCAAGAGTTATAGATTTCCACTGTATTAATCTGGGATACCTATTGAATGTCAG 571
QY	1299	AAGTTCAATTAATGAGCTATTACCTATCAAAAATCAACTGCGCTCAAGGTGGCACTCT 1358
DB	570	AAGTTCAATTAATGAGCTATTACCTATCAAAAATCAACTGCGCTCAAGGTGGCACTCT 511
QY	1359	TTTGTAACAATTTATCCGGAGACACTTTCGACACTGGAGAGAACTGTGGTGGTCAGTT 1418
DB	510	TTTGTAACAATTTATCCGGAGACACTTTCGACACTGGAGAGAACTGTGGTGGTCAGTT 451
QY	1419	ATGGGTTTCTCTGAAAATGCACTGTGACACAGCTGCTTAACCTAGGAGCAGCTAA 1478
DB	450	ATGGGTTTCTCTGAAAATGCACTGTGACACAGCTGCTTAACCTAGGAGCAGCTAA 391
QY	1479	TCAGTGACCACTAGATCTTGCTGCTGTCGCAAAAAGTTATACATCCATCTTTTGAAGCAGT 1538
DB	390	TCAGTGACCACTAGATCTTGCTGCTGTCGCAAAAAGTTATACATCCATCTTTGAAGCAGT 331
QY	1539	TTCTCAATATGCATCGTATTTATGCCATAGACAGAAATGGAGTGACTCTGATGGAAACCCC 1598
DB	330	TTCTCAATATGCATCGTATTTATGCCATAGACAGAAATGGAGTGACTCTGATGGAAACCCC 271
QY	1599	TGACTAAATGACAGGACACAGCCTATTGGATCCCTAGGCTCAATATGCTCCTAATAGTT 1658
DB	270	TGACTAAATGACAGGACACAGCCTATTGGATCCCTAGGCTCAATATGCTCCTAATAGTT 211
QY	1659	ATGGAGTGTGAAGATGTGCACTGCAAAAATTTGGCTCCCAACATTTTTTTTCATCTTA 1718
DB	210	ATGGAGTGTGAAGATGTGCACTGCAAAAATTTGGCTCCCAACATTTTTTTTCATCTTA 151
QY	1719	TTCTTAATGAAGCAACAATCTGAGCTTGCTTCTCTGCAACAGAAATGACACTGCTCTGGG 1778
DB	150	TTCTTAATGAAGCAACAATCTGAGCTTGCTTCTCTGCAACAGAAATGACACTGCTCTGGG 91
QY	1779	ATGAGCTGAGCTCACGTGTGACGCGCACTGATGCTGCGATTAACACTCAAGGAC 1838
DB	90	ATGAGCTGAGGCTCACGTGTGACGCGCACTGATGCTGCGATTAACACTCAAGGAC 31
QY	1839	CATGGGCTTTGTTGGAGGTGGCTGTACT 1867

DB	30	CATGGCTTGTGGAGGGGCTGACTAACT 2
RESULT 15		
AL532425/c		
LOCUS		1013 bp mRNA linear EST 24-MAR-2004
DEFINITION		AL532425 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
ACCESSION		CS0DM014YF11 3-PRIME, mRNA sequence.
VERSION		AL532425
KEYWORDS		AL532425.3 GI:45707342
SOURCE		EST.
ORGANISM		Homo sapiens (human)
REFERENCE		Homo sapiens
AUTHORS		Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL		1 (bases 1 to 1013)
COMMENT		Li, W.B., Gruber, C., Jessee, J. and Polayes, D. Full-length cDNA libraries and normalization Unpublished (2001) On Feb 13, 2001 this sequence version replaced gi:31070257. Contact: Genoscope Genoscope - Centre National de Sequencage 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 9906.r For more information about this cluster, see http://www.genoscope.cns.fr/cdna?s=CS0DM014CC06NP16c=9906.r. Location/Qualifiers 1..1013 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CS0DM014YF11" /tissue_type="FETAL LIVER" /dev_stage="fetal" /note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."
FEATURES		
source		
ORIGIN		
Query Match 29.9%; Score 738; DB 1; Length 1013;		
Best Local Similarity 90.1%; Pred. No. 6.3e-140;		
Matches 798; Conservative 35; Mismatches 48; Indels 5; Gaps 5;		
QY	1432	TGAAATGCACTTTGGACCACTGCTTAACCTAGGAAGCAGCTAATCAGTGACCACGT 1491
DB	938	TGAAATGCACTT-TGACCGCTGCTT-ACCTAKGAAGCGCAGWARTCAGTGMCCACT 881
QY	1492	AGATCTTGTCTGTGCCAAAAAGTTATACATCCATCTTTGAAGCAGTTTCTCA-ATATGC 1550
DB	880	AGATCTTGTCTGTGCCAAAAAGATATACATCCATCTATGANMGCACTTCTCACAATATGC 821
QY	1551	ATCGTATTATTGCCATAGACAGAAATGGAGTGCACTCTGATGGAACCCCTGACTTAAATGA 1610
DB	820	ATCGTATTATTGCCATAGACAGAAATGGAGTGCACTCTGATGGAACCCCTGACTTAAATGA 761
QY	1611	CAGGAACACAGCCTATTGGATCCCTAGGCTCAATATGCTCCTAATAGTTATGGAAGTGA 1670
DB	760	CAGGAACACAGCCTATTGGATCCCTAGGCTCAATATGCTCCTAATAGTTATGGAAGTGA 702
QY	1671	AGATGTGTGCACTGCAAAATTTGGCTCCCAACATTTTTTTCATCTTATCTTAATGAAG 1730
DB	701	AAGAAGTGTGAACCTKCCCTCTATGGCTCCCACTCTCTTTCTCTCTCMACCTACTGCMG 642
QY	1731	CAACAATCTGCAGCTTGCTTCTCTGCAACAGAAATGACACTGCCTGGGATGAGCTGAAGC 1790

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 06:59:18 ; Search time 10581 Seconds
(without alignments)
11320.413 Million cell updates/sec

Title: US-10-616-263-29
Perfect score: 2472
Sequence: 1 tatgagccttcggaacttgt.....aaaaaaaaaaaaaaaaaaaaa 2472

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2472	100.0	2472	6	BD227254	Secreted
2	2429.2	98.3	2497	9	BC028973	Homo sapi
3	2412.2	97.6	2678	9	AF221993	Homo sapi
4	2412	97.6	2512	9	HSM802462	Homo sapi
5	2401	97.1	2516	9	AF221992	Homo sapi
6	2398.4	97.0	2503	9	AK025741	Homo sapi
7	1763.2	71.3	1921	6	BD276479	143 Human
8	1708.2	69.1	1713	6	CQ729071	Sequence
9	1571	63.6	2972	10	BC083863	Rattus no
10	1547.4	62.6	2967	10	BC080765	Mus muscu
11	1545.8	62.5	2699	10	AF254074	Mus muscu
12	1405.6	56.9	74813	9	HS931K24	Human DNA
13	1253.8	50.7	1274	6	BD276497	143 Human
14	908.2	36.7	207304	2	AC107146	Rattus no
15	908.2	36.7	259468	2	AC096067	Rattus no
16	875.6	35.4	195301	10	AL731706	Mouse DNA
17	627	25.4	628	9	HUMY132F03	Homo sapi
18	587.4	23.8	679	9	AF275813	Homo sapi
19	577.8	23.4	1173	10	BC024359	Mus muscu

C	20	457.4	18.5	494	11	G60203	G60203 SHGC-131114
C	21	446.4	18.1	491	6	CQ690654	Sequence
C	22	422.4	17.1	458	6	BD139286	Extended
C	23	413.4	16.7	421	6	BD265148	Compounds
C	24	413.4	16.7	421	6	AR401134	Sequence
C	25	413.4	16.7	421	6	AX192580	Sequence
C	26	371.2	15.0	1831	5	BC045401	Danio rer
C	27	344.6	13.9	348	6	AX887034	Sequence
C	28	344.6	13.9	348	6	BD026644	Sequence
C	29	327.8	13.3	371	6	BD058768	Secreted
C	30	320.4	13.0	428	11	BV103260	MARC 1718
C	31	311.2	12.6	956	10	BC058174	Mus muscu
C	32	279.4	11.3	412	11	BV093483	RPAMMSEQO
C	33	270.8	11.0	378	11	BV093401	RPAMMSEQO
C	34	269	10.9	332	6	AX261806	Sequence
C	35	266.2	10.8	387	11	BV093484	RPAMMSEQO
C	36	263.2	10.6	375	11	BV093400	RPAMMSEQO
C	37	262.4	10.6	390	11	BV093485	RPAMMSEQO
C	38	262.4	10.6	390	11	BV159500	RPAMMSEQO
C	39	260	10.5	379	11	BV093486	RPAMMSEQO
C	40	260	10.5	379	11	BV157816	RPAMMSEQO
C	41	247.6	10.0	1093	5	EX930888	Gallus ga
C	42	246.4	10.0	1099	5	EX934020	Gallus ga
C	43	214.6	8.7	295	11	BV093490	RPAMMSEQO
C	44	214	8.7	291	11	BV157818	RPAMMSEQO
C	45	205.8	8.3	385	11	BV093388	RPAMMSEQO

ALIGNMENTS

RESULT 1	BD227254	2472 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	Secreted proteins and polynucleotides encoding them.				
DEFINITION	BD227254				
ACCESSION	BD227254.1	GI:33037024			
VERSION	JP 2002522062-A/15.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 2472)				
AUTHORS	Jacobs, K., McCoy, J.M., Lavallie, E.R., Racie, L.A.C., Evans, C., Merberg, D., Treacy, M., Agostino, M.J., Ii, R.J.S., Spaulding, V., Wong, G.G., Clark, H.F. and Fechtel, K.				
TITLE	Secreted proteins and polynucleotides encoding them				
JOURNAL	Patent: JP 2002522062-A 15 23-JUL-2002;				
COMMENT	GENETICS INSTITUTE INC				
	OS Homo sapiens (human)				
	PN JP 2002522062-A/15				
	PD 23-JUL-2002				
	PF 13-AUG-1999 JP 2000565001				
	PR 14-AUG-1998 US 60/096622,17-AUG-1998 US 60/096815 PR				
	04-SEP-1998 US 60/099229,23-OCT-1998 US 60/105368 PR				
	08-JAN-1999 US 60/115234,12-FEB-1999 US 60/119931 PR				
	18-FEB-1999 US 60/120575,30-APR-1999 US 60/132020 PR				
	11-AUG-1999 US 60/148424				
	PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A COLLINS				
	PI RACIE,				
	PI CHERYL EVANS, DAVID MERBERG, MAURICE TREACY, MICHAEL J AGOSTINO,				
	PI ROBERT J STEININGER II, VIKKI SPAULDING, GORDON G WONG, HILARY F CLARK,				
	PI KIM FECHTEL				
	PC C12N15/09, A61K38/00, A61K39/395, A61K45/00, A61K48/00,				
	PC A61P7/00,				
	PC A61P7/02, A61P7/04, A61P7/06, A61P13/00, A61P29/00, A61P35/00, PC				
	AC1P37/02,				
	PC A61P43/00, A61P43/00, C07K14/47, C12N5/10, C12P21/02, G01N33/15, PC				
	G01N33/50,				
	PC C12N15/00, C12N5/00, A61K37/02				
	CC Secreted proteins and polynucleotides encoding them FH Key				
	CC Location/Qualifiers				

FEATURES		FT	source	1..2472	/organism='Homo sapiens (human)'	Location/Qualifiers	1..2472
source		FT					
ORIGIN							
Query Match		100.0%; Score 2472; DB 6; Length 2472;					
Best Local Similarity		100.0%; Pred. No. 0;					
Matches 2472; Conservative		0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	TATGAGCCTTCGGAACTTGTGGAGAGACTACAAAGTTTGGTGTGTTATGTCCTTTAGT	60				
DB	1	TATGAGCCTTCGGAACTTGTGGAGAGACTACAAAGTTTGGTGTGTTATGTCCTTTAGT	60				
QY	61	TGGGCTCATACATTTGGGGTGGTACAGAATCAAAAGCAGCCCTGTTTCCAAATACCTAA	120				
DB	61	TGGGCTCATACATTTGGGGTGGTACAGAATCAAAAGCAGCCCTGTTTCCAAATACCTAA	120				
QY	121	AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTCAAATCTTCAGAAGAGCCAAAT	180				
DB	121	AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTCAAATCTTCAGAAGAGCCAAAT	180				
QY	181	CCAGGGGAAGTAGCAGGCTTGCATCTTCAGGTAAAGAGCAGCTTTGAAATCTGAGCTTC	240				
DB	181	CCAGGGGAAGTAGCAGGCTTGCATCTTCAGGTAAAGAGCAGCTTTGAAATCTGAGCTTC	240				
QY	241	ATATCGAAAGAGAGATGAAAAATACAGTTTGGATTAGAAAGAACTGGCTTCTTGTAGCT	300				
DB	241	ATATCGAAAGAGAGATGAAAAATACAGTTTGGATTAGAAAGAACTGGCTTCTTGTAGCT	300				
QY	301	GGGATATCTTTTCATAGGTGTCCATCTTTGGAACATACCTTTTGCAGAGGTCGCAAGCAG	360				
DB	301	GGGATATCTTTTCATAGGTGTCCATCTTTGGAACATACCTTTTGCAGAGGTCGCAAGCAG	360				
QY	361	TCGTGAAAAATTCAGTCTCAAGCAAAACAAAGAGATTGGAAGAGTGAATTAATAAT	420				
DB	361	TCGTGAAAAATTCAGTCTCAAGCAAAACAAAGAGATTGGAAGAGTGAATTAATAAT	420				
QY	421	ATTTGGAAATTAATAATTTGTCTAAATCAATCTATGCTGATTAGCTTCATAAACATTGA	480				
DB	421	ATTTGGAAATTAATAATTTGTCTAAATCAATCTATGCTGATTAGCTTCATAAACATTGA	480				
QY	481	ACTTTTTCATTTTATAGCCACAATGCTGCATATTCATCTTTAAATCTCTAAAGATAAT	540				
DB	481	ACTTTTTCATTTTATAGCCACAATGCTGCATATTCATCTTTAAATCTCTAAAGATAAT	540				
QY	541	TTTAAATGTTAAACGTTGAATGCAATAATAGAAAATGTGGTTTACAAAATAAACCG	600				
DB	541	TTTAAATGTTAAACGTTGAATGCAATAATAGAAAATGTGGTTTACAAAATAAACCG	600				
QY	601	GTCTTCACCTAGTTTACCACTGAAGTAAAGATGTCTCGTTTGGAGCTGAAGAGCCATCAT	660				
DB	601	GTCTTCACCTAGTTTACCACTGAAGTAAAGATGTCTCGTTTGGAGCTGAAGAGCCATCAT	660				
QY	661	GTGTAAGAGTGAAACCACTGACAACTGAGAGAGTCAGGACCAACATTTCTGCTTGAAG	720				
DB	661	GTGTAAGAGTGAAACCACTGACAACTGAGAGAGTCAGGACCAACATTTCTGCTTGAAG	720				
QY	721	AATTGTAAATCATCATGCTATGGCCCTTCAGGTAGGCTGAAGCAGCTGCACAATGGCTTGG	780				
DB	721	AATTGTAAATCATCATGCTATGGCCCTTCAGGTAGGCTGAAGCAGCTGCACAATGGCTTGG	780				
QY	781	AGGTTACGTTGTGTACAACTTCACAGTCTTCAGTCTGCTCAGTCACCTTTTGGTTCACACA	840				
DB	781	AGGTTACGTTGTGTACAACTTCACAGTCTTCAGTCTGCTCAGTCACCTTTTGGTTCACACA	840				
QY	841	TCCCATTTTAAAGATCCTGACAGCCCTCCATACAGAATCATGTGTCAAGCTTCAGTGATTG	900				
DB	841	TCCCATTTTAAAGATCCTGACAGCCCTCCATACAGAATCATGTGTCAAGCTTCAGTGATTG	900				

901	QY	TGGCTTATTCACAGCTATTCTTTGCTGCAACCTGATTGAAAATGTTTCAGAGATTAGCCTT	960
901	DB	TGGCTTATTCACAGCTATTCTTTGCTGCAACCTGATTGAAAATGTTTCAGAGATTAGCCTT	960
961	QY	GACACCCACCACTGCTATTAGATTAAATAAAACATCTTTTTCAGTCTTTTGCATCAGTTATCT	1020
961	DB	GACACCCACCACTGCTATTAGATTAAATAAAACATCTTTTTCAGTCTTTTGCATCAGTTATCT	1020
1021	QY	CAAGTCTGATACCTGTGCTGGAATCCCAAGTGGACTTTTGTAGTAGTACTCAGATCCCTCCT	1080
1021	DB	CAAGTCTGATACCTGTGCTGGAATCCCAAGTGGACTTTTGTAGTAGTACTCAGATCCCTCCT	1080
1081	QY	TTGTTTGGTGGTAGTATATTAAACAAGTAAACCTTGCTGTATGCTCACCAAGAAGGAAAC	1140
1081	DB	TTGTTTGGTGGTAGTATATTAAACAAGTAAACCTTGCTGTATGCTCACCAAGAAGGAAAC	1140
1141	QY	AGAGCATGTCAAGTCTTTTGATCTCTGAGAGCCTTTTGTCTTACAATTCAGAAAAATGCTGA	1200
1141	DB	AGAGCATGTCAAGTCTTTTGATCTCTGAGAGCCTTTTGTCTTACAATTCAGAAAAATGCTGA	1200
1201	QY	AGGCCACATCATTTTGTAGGAAAGAGTTTAAATTGTACTTTTAAAGAGTCAAGAGCTTATAGA	1260
1201	DB	AGGCCACATCATTTTGTAGGAAAGAGTTTAAATTGTACTTTTAAAGAGTCAAGAGCTTATAGA	1260
1261	QY	TTCCACTGTATTACTCGGGTACTCATTTGAAATGTTCAGAAATGTTCAATTAATAGGCTATT	1320
1261	DB	TTCCACTGTATTACTCGGGTACTCATTTGAAATGTTCAGAAATGTTCAATTAATAGGCTATT	1320
1321	QY	ACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCCGAGA	1380
1321	DB	ACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCCGAGA	1380
1381	QY	CACTTCTGACACTGCGAAGGAACTGTGTGGTCAGTTATGGGGTTTCTCTCAAAAATGTC	1440
1381	DB	CACTTCTGACACTGCGAAGGAACTGTGTGGTCAGTTATGGGGTTTCTCTGAAAATGTC	1440
1441	QY	AGTCTTGGACCACTGCTTTAACTTAGAAGGCGAGCTTAATCAGTGAACAAGTAGATCTTGT	1500
1441	DB	AGTCTTGGACCACTGCTTTAACTTAGAAGGCGAGCTTAATCAGTGAACAAGTAGATCTTGT	1500
1501	QY	CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGCATCGTATTAT	1560
1501	DB	CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGCATCGTATTAT	1560
1561	QY	TGCCATAGACAGAAATGGAGTGACTCTGTATGGAAACCCCTGACATAAAATGACAGGAAACA	1620
1561	DB	TGCCATAGACAGAAATGGAGTGACTCTGTATGGAAACCCCTGACATAAAATGACAGGAAACA	1620
1621	QY	GCCTATTGGATCCCTAGGCTCAATATGCTCCTTAATAGTTATGGAAGTGTGAAAGATGTTG	1680
1621	DB	GCCTATTGGATCCCTAGGCTCAATATGCTCCTTAATAGTTATGGAAGTGTGAAAGATGTTG	1680
1681	QY	CACGTGCAAAATTTGGCTCCCAACATTTTTTTTTCATCTTATTCTTAATGAAGCAACAATCTG	1740
1681	DB	CACGTGCAAAATTTGGCTCCCAACATTTTTTTTTCATCTTATTCTTAATGAAGCAACAATCTG	1740
1741	QY	CAGCTTGTCTCTGCAACAGAAATGACACTGCTCGGATGAGCTGAAGCTCACGTGTCA	1800
1741	DB	CAGCTTGTCTCTGCAACAGAAATGACACTGCTCGGATGAGCTGAAGCTCACGTGTCA	1800
1801	QY	GACGCACTGCACTGCTGCAAGTTAAACACTCAAGGAACCATGGCTTTGTTGGAGGTGG	1860
1801	DB	GACGCACTGCACTGCTGCAAGTTAAACACTCAAGGAACCATGGCTTTGTTGGAGGTGG	1860
1861	QY	CTGTACTGAAACTCAITTTGGCTGCATATATCAGACAAAGACTCACAAACGCCAGAAAG	1920
1861	DB	CTGTACTGAAACTCAITTTGGCTGCATATATCAGACAAAGACTCACAAACGCCAGAAAG	1920
1921	QY	CATTCTCAAGATGATGATGACTCAACAGAACTTCAATTAATGCTGAAGCATTTTG	1980
1921	DB	CATTCTCAAGATGATGATGACTCAACAGAACTTCAATTAATGCTGAAGCATTTTG	1980
1981	QY	CAGTGCCTCAGAATCTGTGTTGGCTCTTTTAGAACAATGATGGAGGTGAAATTTCTCACTGA	2040

QY	901	TGGCTTATTCACAGCTATTTCTTGTGCAACCTGATGAAATGTTTCAGAGATTAGGCTT	960				
DB	901	TGGCTTATTCACAGCTATTTCTTGTGCAACCTGATGAAATGTTTCAGAGATTAGGCTT	960				
QY	961	GACACCCACCACCTGTCTATTAGATTAAATAAAACATCTTTTGTAGCTTTTGTGATCTATCT	1020				
DB	961	GACACCCACCACCTGTCTATTAGATTAAATAAAACATCTTTTGTAGCTTTTGTGATCTATCT	1020				
QY	1021	CAAGTCTGANTACCTGTGGTTGTCGAATCCAGTGGACTTTTAGTAGTACTCAGATCCTCT	1080				
DB	1021	CAAGTCTGANTACCTGTGGTTGTCGAATCCAGTGGACTTTTAGTAGTACTCAGATCCTCT	1080				
QY	1081	TTGTTTGGTGGCTAGTATATTAAACAAGTAAACCTGCTGTATGCTCACACAGAAAGAAAC	1140				
DB	1081	TTGTTTGGTGGCTAGTATATTAAACAAGTAAACCTGCTGTATGCTCACACAGAAAGAAAC	1140				
QY	1141	AGAGCATGTCACTGTCTTTGATCCTCGAGAGCCCTTTTGTCTTCAATTCAGAAAAATGCTGA	1200				
DB	1141	AGAGCATGTCACTGTCTTTGATCCTCGAGAGCCCTTTTGTCTTCAATTCAGAAAAATGCTGA	1200				
QY	1201	AGGCCACATCATTTTAGGAAAGAGTTTAATTTGTACTCTTTTAAAGGTCAAAGATTATAGA	1260				
DB	1201	AGGCCACATCATTTTAGGAAAGAGTTTAATTTGTACTCTTTTAAAGGTCAAAGATTATAGA	1260				
QY	1261	TTTCCACTGTATTTACCTGGGATACTCAATTGAAATGTGAGAAGTTCAATTAATGAGGCTATT	1320				
DB	1261	TTTCCACTGTATTTACCTGGGATACTCAATTGAAATGTGAGAAGTTCAATTAATGAGGCTATT	1320				
QY	1321	ACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACCTTTATCCGGAGA	1380				
DB	1321	ACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACCTTTATCCGGAGA	1380				
QY	1381	CACTTCTGACACTGAGAAAGAACTGTGGTGTGCTAGTTATGGGGTTTCTCTTGAATAATGC	1440				
DB	1381	CACTTCTGACACTGAGAAAGAACTGTGGTGTGCTAGTTATGGGGTTTCTCTTGAATAATGC	1440				
QY	1441	AGCTTGGACCAAGCTGCTTAACTAGGAGGAGCTTAATCACTGACCACTGATAGTCTTCT	1500				
DB	1441	AGCTTGGACCAAGCTGCTTAACTAGGAGGAGCTTAATCACTGACCACTGATAGTCTTCT	1500				
QY	1501	CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGSCATCGTATTAT	1560				
DB	1501	CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGSCATCGTATTAT	1560				
QY	1561	TGCCATAGACAGAAATTTGGAGTACTCTGATGGAACCCCTGACTTAAATGACAGGACACA	1620				
DB	1561	TGCCATAGACAGAAATTTGGAGTACTCTGATGGAACCCCTGACTTAAATGACAGGACACA	1620				
QY	1621	GCCTATTGGATCCCTAGGCTCAATATGCTTAATAGTTATGGAAGTGTGAAAGATGTGTG	1680				
DB	1621	GCCTATTGGATCCCTAGGCTCAATATGCTTAATAGTTATGGAAGTGTGAAAGATGTGTG	1680				
QY	1681	CACTGCAAAATTTGGCTCCCAACATTTTTCATCTTATTCCTAATGAAGCAACAATCTG	1740				
DB	1681	CACTGCAAAATTTGGCTCCCAACATTTTTCATCTTATTCCTAATGAAGCAACAATCTG	1740				
QY	1741	GAGCTTGTCTTCTGCAACAGAAATGACACTGCTGCTGGGATGAGCTGAGCTCACGTGTCA	1800				
DB	1741	GAGCTTGTCTTCTGCAACAGAAATGACACTGCTGCTGGGATGAGCTGAGCTCACGTGTCA	1800				
QY	1801	GAGGCACTGCAATGCTGCACTTAACACTCAAGAACCCCTGGCTTTTCTGGGAGGTGG	1860				
DB	1801	GAGGCACTGCAATGCTGCACTTAACACTCAAGAACCCCTGGCTTTTCTGGGAGGTGG	1860				
QY	1861	CTGTACTGAAACTCAATTTGGCTGCATATATACAGACTCAAGACTCAACAGCCAGAAAG	1920				
DB	1861	CTGTACTGAAACTCAATTTGGCTGCATATATACAGACTCAAGACTCAACAGCCAGAAAG	1920				
QY	1921	CAATCTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1980				
DB	1921	CAATCTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1980				
QY	1981	CAGTGCCCTTAGAATCTGTGTTGGTCTTTTAGAAATCATGATGGAGGTGAAATTTCTCACTGA	2040				

QY 121 AAACGACGACATTCCTGACGACAGATAGTCTGGACCTTTCARATCTTCAGAAAGCCCAAT 180
DB 176 AAACGACGACATTCCTGACGACAGATAGTCTGGACCTTTCARATCTTCAGAAAGCCCAAT 235
QY 181 CCAGGGGAAGTAGCAGGCTTGCATCTTCAGGTAAAGAACGACGCTTGAATCTGAGCTTC 240
DB 236 CCAGGGGAAGTAGCAGGCTTGCATCTTCAGGTAAAGAACGACGCTTGAATCTGAGCTTC 295
QY 241 ATATCGAAAGAGAGATGAAATAATACAGTGTGAATAGAAAGAACTGGCTTCTTGTAGCT 300
DB 296 ATATCGAAAGAGAGATGAAATAATACAGTGTGAATAGAAAGAACTGGCTTCTTGTAGCT 355
QY 301 GGGATATCTTTCATAGGTCTCCATCTTGGACATATCTTTTCGACAGGCTTCGAAAGCAG 360
DB 356 GGGATATCTTTCATAGGTCTCCATCTTGGACATATCTTTTCGACAGGCTTCGAAAGCAG 415
QY 361 TCTGTAAATTTTCAGTCTCAAGCAAAACAAAGAGATTTGAAGAGTGAAGTAAATAAAT 420
DB 416 TCTGTAAATTTTCAGTCTCAAGCAAAACAAAGAGATTTGAAGAGTGAAGTAAATAAAT 475
QY 421 ATTTGGAAATTAATTTGTCTATAAATCATTTCTATGCTGATPAGCTTCATAAACAATTGA 480
DB 476 ATTTGGAAATTAATTTGTCTATAAATCATTTCTATGCTGATPAGCTTCATAAACAATTGA 535
QY 481 ACTTTTGTATTTATAGCCACAAATGCTGCATATTCATCTTTAAATTCCTTAAAGATAAT 540
DB 536 ACTTTTGTATTTATAGCCACAAATGCTGCATATTCATCTTTAAATTCCTTAAAGATAAT 595
QY 541 TTTAAATGTTAAACGCTGATAATGCAATAAATAGAAAAATGTGGTTTACAAAAATAAAACG 600
DB 596 TTTAAATGTTAAACGCTGATAATGCAATAAATAGAAAAATGTGGTTTACAAAAATAAAACG 655
QY 601 GTCTTCACCTAGTTPACCACTGAAGTAGATGTCTCGTTTGGAAAGCTPAGAAGCCATCAT 660
DB 656 GTCTTCACCTAGTTPACCACTGAAGTAGATGTCTCGTTTGGAAAGCTPAGAAGCCATCAT 715
QY 661 GTGTAGAGTGAACCACTGACAACTGAGAGAGTCAAGGACACATTTCTGTCTTGAAGA 720
DB 716 GTGTAGAGTGAACCACTGACAACTGAGAGAGTCAAGGACACATTTCTGTCTTGAAGA 775
QY 721 AATTGTAAACATCATGTATGGCCCTTCAGTAGGCTGGAAGCAGCTGCACAAATGGCTTTGG 780
DB 776 AATTGTAAACATCATGTATGGCCCTTCAGTAGGCTGGAAGCAGCTGCACAAATGGCTTTGG 835
QY 781 AGTTACGTGTGTPAACCTCAAGTCTCAAGTCTGCTGCTCAGTCACTTTTGGTCAACA 840
DB 836 AGTTACGTGTGTPAACCTCAAGTCTCAAGTCTGCTGCTCAGTCACTTTTGGTCAACA 895
QY 841 TCCCATTTTAAAGATCCTGACAGCCCTCCATACAGATCATGTGTCAAGCTTCAGTGATTG 900
DB 896 TCCCATTTTAAAGATCCTGACAGCCCTCCATACAGATCATGTGTCAAGCTTCAGTGATTG 955
QY 901 TGGCTTTATTCACAGCTATTTCTTGTGCAACCTGATTGAAAAATGTTTCAGAGATTAGGCTT 960
DB 956 TGGCTTTATTCACAGCTATTTCTTGTGCAACCTGATTGAAAAATGTTTCAGAGATTAGGCTT 1015
QY 961 GACACCCAACCTGTCAATAGATTAAATAAACAATCTTTTGAAGTCTTTTGCATCAGTTATCT 1020
DB 1016 GACACCCAACCTGTCAATAGATTAAATAAACAATCTTTTGAAGTCTTTTGCATCAGTTATCT 1075
QY 1021 CAAGTCTGATACCTGTGGTGTGCAATCCAGTGGACCTTTAGTAGTACTCAGATCCTCCT 1080
DB 1076 CAAGTCTGAGACCTGTGGTGTGCAATCCAGTGGACCTTTAGTAGTACTCAGATCCTCCT 1135
QY 1081 TTGTTTGGTGGCTAGTATATTAACAAGTAAACCTGCTGTATGCTCACCCAGAAAGGAAC 1140
DB 1136 TTGTTTGGTGGCTAGTATATTAACAAGTAAACCTGCTGTATGCTCACCCAGAAAGGAAC 1195
QY 1141 AGAGCATGTCAAGTCTTGTGATCCTGAGAGCTTTTGTCTTCAATTCAGAAATAAGTCTGA 1200
DB 1196 AGAGCATGTCAAGTCTTGTGATCCTGAGAGCTTTTGTCTTCAATTCAGAAATAAGTCTGA 1255

QY 1201 AGGCCACACATCATTTTAGGAAAGAGTTTAATTCGTCTTAAAGGTCAAAGAGTTATAGA 1260
DB 1256 AGGCCACACATCATTTTAGGAAAGAGTTTAATTCGTCTTAAAGGTCAAAGAGTTATAGA 1315
QY 1261 TTCCACTGTATTTACCTGGGATACCTCATTTGAATGTGAGAAAGTTCAATTAATGAGGCTATT 1320
DB 1316 TTCCACTGTATTTACCTGGGATACCTCATTTGAATGTGAGAAAGTTCAATTAATGAGGCTATT 1375
QY 1321 ACCTATCAAAAAAATCAACTGCCCTCAAGGTGGACCTCTTTTGTACAACCTTTATCCGGAGA 1380
DB 1376 ACCTATCAAAAAAATCAACTGCCCTCAAGGTGGACCTCTTTTGTACAACCTTTATCCGGAGA 1435
QY 1381 CACTTCTGCACACTGGAGAGGAACCTGCTGTGTGCTCAGTTATGCGGTTTCTCTTGAATAATGC 1440
DB 1436 CACTTCTGCACACTGGAGAGGAACCTGCTGTGTGCTCAGTTATGCGGTTTCTCTTGAATAATGC 1495
QY 1441 AGCTTTGACCACTGCTTAACTAGGAAGGAGCTAATCAGTGACCACCTAGATCTTTGT 1500
DB 1496 AGCTTTGACCACTGCTTAACTAGGAAGGAGCTAATCAGTGACCACCTAGATCTTTGT 1555
QY 1501 CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGCTATGCTATTAT 1560
DB 1556 CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGCTATGCTATTAT 1615
QY 1561 TGCCATAGACAGAAATTTGGAGTGACCTCTGATGGAAACCCCTGACTAAAATGACAGGAACACA 1620
DB 1616 TGCCATAGACAGAAATTTGGAGTGACCTCTGATGGAAACCCCTGACTAAAATGACAGGAACACA 1675
QY 1621 GCCTATTGGATCCCTAGGCTCAATATGCTCTAATAGTTATGGAAGTGTGAAAGATGTGTG 1680
DB 1676 GCCTATTGGATCCCTAGGCTCAATATGCTCTAATAGTTATGGAAGTGTGAAAGATGTGTG 1735
QY 1681 CACTGCAAAAATTTGGCTCCCAACATTTTTCATCTTATTCCTAATGAAGCAACAATCTG 1740
DB 1736 CACTGCAAAAATTTGGCTCCCAACATTTTTCATCTTATTCCTAATGAAGCAACAATCTG 1795
QY 1741 CAGCTTGTCTCTGCAACAGAAATGACACTGSCCTGGGATGAGTGAAAGTCACTCAGCTGTCA 1800
DB 1796 CAGCTTGTCTCTGCAACAGAAATGACACTGSCCTGGGATGAGTGAAAGTCACTCAGCTGTCA 1855
QY 1801 GACGCACTGCAATGCTCTGCACTTAACACTCAAGGAACCATGGGCTTTGTTGGGAGGTGG 1860
DB 1856 GACGCACTGCAATGCTCTGCACTTAACACTCAAGGAACCATGGGCTTTGTTGGGAGGTGG 1915
QY 1861 CTGTACTGAACCTCATTTGGCTGCAATATACAGACAAGACTCAACGACCCAGAAAG 1920
DB 1916 CTGTACTGAACCTCATTTGGCTGCAATATACAGACAAGACTCAACGACCCAGAAAG 1975
QY 1921 CATTCTCAAGATGATGATGACTCAACAGAACTTCAATTTAATTTGCTGAAGCATTTTG 1980
DB 1976 CATTCTCAAGATGATGATGACTCAACAGAACTTCAATTTAATTTGCTGAAGCATTTTG 2035
QY 1981 CAGTGCCCTTAGAATCTGTTTGGCTCTTTTAGAACATGATGAGGTGAAATTTCTCACTGA 2040
DB 2036 CAGTGCCCTTAGAATCTGTTTGGCTCTTTTAGAACATGATGAGGTGAAATTTCTCACTGA 2095
QY 2041 CATGAAGTATGACACCTTTTGGTCAAGTTCAGGCAAGTTCCTCCTGTGTTGCTAACTGGCC 2100
DB 2096 CATGAAGTATGACACCTTTTGGTCAAGTTCAGGCAAGTTCCTCCTGTGTTGCTAACTGGCC 2155
QY 2101 AGATTTTGTCTTACAGAGTGGCTGGGATTAACAAATGACAGGAAGAACTCAACTGGTC 2160
DB 2156 AGATTTTGTCTTACAGAGTGGCTGGGATTAACAAATGACAGGAAGAACTCAACTGGTC 2215
QY 2161 TTTCTTAAGAGACACACGCTCGTCCATTTTGTGCAAAAAGTGCCTTTCCAATGAAAGCTGT 2220
DB 2216 TTTCTTAAGAGACACATGCTGCTCCATTTTGTGCAAAAAGTGCCTTTCCAATGAAAGCTGT 2275
QY 2221 GGGCTCAGCAGCAACCTGACCTTGGACTGTTTGTGCTGAAAGCTTAGTGGCTCAGGT 2280
DB 2276 GGTCTCAGCAGCAACCTGACCTTGGACTGTTTGTGCTGAAAGCTTAGTGGCTCAGGT 2335
QY 2281 GGCTGTAGAGACAGCAATTTTGATTTGGATCTTTTCATATGTTATTGAAGATAAAAACTA 2340

Db 2336 GGCCTGTAGACACAGCCAAATTTGGATCTTTCATATGTTATTGAAGATAAAACTA 2395
Qy 2341 AGAAGATAGCATGTTCTGATTACAGAGAGAAACAAATAAACTAGTCTCTTGGCAATTGAGA 2400
Db 2396 AGAGATAGCATGTTCTGATTACAGAGAGAAACAAATAAACTAGTCTCTTGGCAATTGAAA 2455
Qy 2401 AA 2442
Db 2456 AA 2497

RESULT 3
AF221993
LOCUS 2678 bp mRNA linear PRI 02-JUN-2000
DEFINITION Homo sapiens putative chaperonin MKKS (MKKS) mRNA, complete cds,
alternatively spliced.
ACCESSION AF221993.1 GI:8163833
VERSION 1
KEYWORDS Mutation of a gene encoding a putative chaperonin causes
SOURCE McKusick-Kaufman syndrome
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2678)
AUTHORS Stone,D.L., Slavotinek,A., Bouffard,G.G., Banerjee-Basu,S.,
Baxevarian,A.D., Barr,M. and Biesecker,L.G.
TITLE Mutation of a gene encoding a putative chaperonin causes
McKusick-Kaufman syndrome
JOURNAL Nat. Genet. 25 (1), 79-82 (2000)
MEDLINE 20264373
PUBMED 10802661
REFERENCE 2 (bases 1 to 2678)
AUTHORS Biesecker,L.G., Slavotinek,A.M. and Stone,D.L.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-2000) LGDR, NHGRI/NIH, 49 Convent Drive,
Bethesda, MD 20892-4472, USA

FEATURES
source
1..2678
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="20p12"
/notes="corresponds to genomic sequence deposited in
GenBank Accession Number AL034430"
1..2678
/genes="MKKS"
/notes="McKusick-Kaufman (hydrometrocoplos-polydactyly)
syndrome"
888..2600
/gene="MKKS"
/note="transcript MKSB; includes alternative exon 1B"
/codon_start=1
/product="putative chaperonin MKKS"
/protein_id="AAF73873.1"
/db_xref="GI:8163834"
/translation="MSRLAKKPSLCKSEPLTTERVTTLSVLKRVTSYCSYSGRLK
QLNKGGVCTTSQSALLSHLLVTHPIKLKLTASIQNHVSFSDGCFALLCNL
IENVRLGUTPTTVIRNKHLLSLCISYLKSETCGCRIPVDFSSQIILCLVRSILTS
KPMALTRKETHEVLSALILRAFLFTIPENAEGBHIIILKSLIPLKQRVIDSTVLFGI
LIEMSVQMLRLPIPKKSTALKVAFCTTLSGDTGRTGVVSVYGSLENAVLDPOL
NLGRQLISDHVDLVCQKVIHPSLKQFLNMHRI IADIRIGVTLMPLTKMTGTPIG
SLGSI CPNSYGSVKDVTAKFGSKHFLI PNEATICSILLGNRNDTAWDELKLCQT
ALHVQLITLKEPWALLGGCTETHLAAYIRKHTNDPESILKDECTQELQLIAEAF
CSALESVSGLEHDGGEILTDMKYHLSVQADSPCVANWPDLLSCQCGLYNSQEEL
NWSFLRSTRPFVQPSCLPHEAVGSASNLTDCLTAKLSGLQVAVETANLILDLSYVI
EDKN"

ORIGIN

Query Match 97.6%; Score 2412.2; DB 9; Length 2678;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2414; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1081 TTGTTTGGTGGCTAGTATATTAAACAAGTAAACCTGCTGTATGCTCACCAAGAGGAAAC 1140
|||||
Db 1340 TTGTTTGGTGGCTAGTATATTAAACAAGTAAACCTGCTGTATGCTCACCAAGAGGAAAC 1399
|||||
QY 1141 AGAGCATGTGTCAGTGTCTTGCATCTCTGAGAGCCTTTTGGCTTACAAATTCACAGAAATGCTGA 1200
|||||
Db 1400 AGAGCATGTGTCAGTGTCTTGCATCTCTGAGAGCCTTTTGGCTTACAAATTCACAGAAATGCTGA 1459
|||||
QY 1201 AGGCCACATCATTTTAAAGAAAGTAAATTTGTCCTTTTAAAGGTCACAAAGAGTTATAGA 1260
|||||
Db 1460 AGGCCACATCATTTTAAAGAAAGTAAATTTGTCCTTTTAAAGGTCACAAAGAGTTATAGA 1519
|||||
QY 1261 TTCCACTGTATTACCTGGGATCTCAATTTGAATGTGAGAGTTCAATTTAATGAGGCTATT 1320
|||||
Db 1520 TTCCACTGTATTACCTGGGATCTCAATTTGAATGTGAGAGTTCAATTTAATGAGGCTATT 1579
|||||
QY 1321 ACCTATCAAAAAATCAACTGCCCTCAAGGTGGGCACTTTTGTACAACTTTATCCGGAGA 1380
|||||
Db 1580 ACCTATCAAAAAATCAACTGCCCTCAAGGTGGGCACTTTTGTACAACTTTATCCGGAGA 1639
|||||
QY 1381 CACTTCTGACATGGAAGAGAACTGTGTGTGTCAGTTATGGGTTTCTTTGAAATGTC 1440
|||||
Db 1640 CACTTCTGACATGGAAGAGAACTGTGTGTGTCAGTTATGGGTTTCTTTGAAATGTC 1699
|||||
QY 1441 AGTCTTGGACCACTGCTTAACTAGGAAGGCGAGCTAATCAGTGACCACTAGATCTTGT 1500
|||||
Db 1700 AGTCTTGGACCACTGCTTAACTAGGAAGGCGAGCTAATCAGTGACCACTAGATCTTGT 1759
|||||
QY 1501 CCTGTGCCAAAAAGTTATACATCCATCTTTGAAGCAGTTTCTCAATATGCATCGTATTAT 1560
|||||
Db 1760 CCTGTGCCAAAAAGTTATACATCCATCTTTGAAGCAGTTTCTCAATATGCATCGTATTAT 1819
|||||
QY 1561 TGCATPAGACAGAAATGGAGTGACTCTGTATGGAACCCCTGACTAAATGACAGGAACACA 1620
|||||
Db 1820 TGCATPAGACAGAAATGGAGTGACTCTGTATGGAACCCCTGACTAAATGACAGGAACACA 1879
|||||
QY 1621 GCCTATTGATCCCTAGGCTCAATATGCTTAATAGTTATGGAAGTGAAGATGTGTG 1680
|||||
Db 1880 GCCTATTGATCCCTAGGCTCAATATGCTTAATAGTTATGGAAGTGAAGATGTGTG 1939
|||||
QY 1681 CACTGCAAAATTTGGCTCCCAACATTTTTCATCTTATTCCTAATGAAGCAACAATCTG 1740
|||||
Db 1940 CACTGCAAAATTTGGCTCCCAACATTTTTCATCTTATTCCTAATGAAGCAACAATCTG 1999
|||||
QY 1741 CAGCTTGTCTTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCAGGTGTC 1800
|||||
Db 2000 CAGCTTGTCTTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCAGGTGTC 2059
|||||
QY 1801 GACGGCACTGCATGCTCTGCAGTTAACACTCAAGGAACCATGGGCTTTGTTGGGAGGTGG 1860
|||||
Db 2060 GACGGCACTGCATGCTCTGCAGTTAACACTCAAGGAACCATGGGCTTTGTTGGGAGGTGG 2119
|||||
QY 1861 CTGTACTGAAACTCATTTGGCTGCATATATCAGACACAAGACTCACAAGCCAGCAAAAG 1920
|||||
Db 2120 CTGTACTGAAACTCATTTGGCTGCATATATCAGACACAAGACTCACAAGCCAGCAAAAG 2179
|||||
QY 1921 CATTTCTCAAAGATGATGAATGTAATCAACAGAACTTCAATTAATGCTGAAGCAATTTG 1980
|||||
Db 2180 CATTTCTCAAAGATGATGAATGTAATCAACAGAACTTCAATTAATGCTGAAGCAATTTG 2239
|||||
QY 1981 CAGTGCCTTAGAATCTGTTGGCTCTTTAGAACATGATGAGGTTGAATTTCTCACTGA 2040
|||||
Db 2240 CAGTGCCTTAGAATCTGTTGGCTCTTTAGAACATGATGAGGTTGAATTTCTCACTGA 2299
|||||
QY 2041 CATGAAGATGGAACACTTTGGTCAGTTTCAAGGAGAGATTTCTCCTGTGTTGCTAACTGGCC 2100
|||||
Db 2300 CATGAAGATGGAACACTTTGGTCAGTTTCAAGGAGAGATTTCTCCTGTGTTGCTAACTGGCC 2359
|||||
QY 2101 AGATTTGCTTTCAAGTGTGGCTGTGGATATATCAATAGCAGGAAGAACTCAACTGCTGTC 2160
|||||
Db 2360 AGATTTGCTTTCAAGTGTGGCTGTGGATATATCAATAGCAGGAAGAACTCAACTGCTGTC 2419
|||||
QY 2161 TTTCTTTAAGAGCACACGTCGTCTCATTTGTGCCCACAAAGCTGCTTCCACATGAAGCTGT 2220
|||||

Db 2420 TTTCTTAAAGACACACAGCTGCTCCATTTGTGCCACAAAGCTGCCCTTCCACATGAAGCTGT 2479
|||||
QY 2221 GGGCTCAGCCAGCAACCTGACCTTGGACTGTTTGAAGCTGCAAAAGCTTAGTGGCTACAGGT 2280
|||||
Db 2480 GGGCTCAGCCAGCAACCTGACCTTGGACTGTTTGAAGCTGCAAAAGCTTAGTGGCTACAGGT 2539
|||||
QY 2281 GGCTGTAGAGACAGCCAAATTTGATTGGGATCTTTTCATATGTTATGTAAGATAAAACTA 2340
|||||
Db 2540 GGCTGTAGAGACAGCCAAATTTGATTGGGATCTTTTCATATGTTATGTAAGATAAAACTA 2599
|||||
QY 2341 AGAGAAATGACATGTTTCGTATTACAGAGAAACAAATAAAGCTAGTCTGTTGGCAATTGAGA 2400
|||||
Db 2600 AGAGAAATGACATGTTTCGTATTACAGAGAAACAAATAAAGCTAGTCTGTTGGCAATTGAGA 2659
|||||
QY 2401 AAAAAAAAAAAAAAAAAAAAAA 2419
|||||
Db 2660 AAAAAAAAAAAAAAAAAAAAAA 2678
|||||
RESULT 4
HSM802462
LOCUS HSM802462 2512 bp mRNA linear PRI 18-FEB-2000
DEFINITION Homo sapiens mRNA; cDNA DKFZp761A072 (from clone DKFZp761A072).
ACCESSION AL157427
VERSION AL157427.1 GI:7018456
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2512)
AUTHORS Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2000) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
COMMENT Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by MedGenomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp761A072) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://www.mips.biochem.mpg.de/proj/cDNA/>.
FEATURES
Location/Qualifiers
1..2512
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="32.60 cR from top of Chr20 linkage group"
/clone="DKFZp761A072"
/tissue_type="amygdala"
/clone_lib="761 (synonym: hamy2). Vector pSport1; host DH10B; sites NotI + SalI"
/dev_stage="adult"
1..2512
/gene="DKFZp761A072"
721..2433
/gene="DKFZp761A072"
/note="unknown protein"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB75652.1"
/db_xref="GI:7018457"
/db_xref="GOA:O9NPJ1"
/translation="MSRLAEKKPSLCKSEPLTTERVRTLVLKLVITSCYSPSRLK QLNHFGGVCCTSSSALLSHLVTHPIKLITASIQNHVSSPDSGGLFTAILCCNL IENVQLGLTPTTVIRLNKHLISLCISYKSETCCGRIPVDFPSSQTILLCLVRSILTS KPACMLTKRETHVSALIRAFLLTIPENAEGLIILGKSLIVPLKQGVISLENAVLPGI LIEMSEVOLMRLLPIKSKTALKVALFCTTISGDTSDTGTGTVVWSYGVSLNVLVDQL LNLGRQLISDHVDLVLCQKVIHPSLKQFLNMHRIIADRGVTLMEPLTKMTGTQPIG

SLGICPNSYGSVDVCTAKFGSKPHLIPNEATICSLLLCNRNDPTAWDELKLTQOT
ALHVLQUTLEKPEWALLGGCTETHLAAYIRKTHNDPESILKDDKCTQCELOLIAEAF
CSALESVVGSLRHGGBILDMKYGHLSVQADSPCVANWPDLLSCGCGLYNSOEL
NWSFLRSTRPFPQSCLPHEAVGSASNLTLDCITAKLSGLQVAVETANLILDLISYVI
EDKN"
polyA_signal 2466..2471
/gene="DKFZp761A072"
polyA_site 2489
/gene="DKFZp761A072"
ORIGIN

Query Match 97.6%; Score 2412; DB 9; Length 2512;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2415; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy	1	TATGAGCCTTTCGGAACCTTGTGGAGAGACTACAAAGTTTTCGTTGTATGCTCCCTTTAGT	60
Db	93	TATGAGCCTTCGGAACTTGTGGAGAGACTACAAAGTTTTCGTTGTATGCTCCCTTTAGT	152
Qy	61	TGGGCTCATACATTTTGGGGTGGTACAGAATCAAAAGCAGCCCTGTTTTCCAAATACCTAA	120
Db	153	TGGGCTCATACATTTTGGGGTGGTACAGAATCAAAAGCAGCCCTGTTTTCCAAATACCTAA	212
Qy	121	AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTTCAAATCTTCAGAAGAGCCAAAT	180
Db	213	AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTTCAAATCTTCAGAAGAGCCAAAT	272
Qy	181	CCAGGGGAAGTACGAGCTTGCATCTTCAGGTAAAGAGCAGCTTTGAATCAGGCTTC	240
Db	273	CCAGGGGAAGTACGAGCTTGCATCTTCAGGTAAAGAGCAGCTTTGAATCAGGCTTC	332
Qy	241	ATATCGAAAGAGAGATGAAATAATACCAAGTTGGATTAGAAAGAACTGGCTTCTTGTAGCT	300
Db	333	ATATCGAAAGAGAGATGAAATAATACCAAGTTGGATTAGAAAGAACTGGCTTCTTGTAGCT	392
Qy	301	GGGATATCTTTTCAATAGGTGTCATCTTTGGAAACATACATTTTTCAGAGGCTGCAAAAGCAG	360
Db	393	GGGATATCTTTTCAATAGGTGTCATCTTTGGAAACATACATTTTTCAGAGGCTGCAAAAGCAG	452
Qy	361	TCGTGTAATTTTCAGTCTCAAGCAACAAAGAGATTTGAAGAGTGAAGTAAATTAAT	420
Db	453	TCGTGTAATTTTCAGTCTCAAGCAACAAAGAGATTTGAAGAGTGAAGTAAATTAAT	512
Qy	421	ATTGTGAATTTTAAATTTGTCAATTAATCATCTATGCTGATTAAGTCTTCATAAACAATTGA	480
Db	513	ATTGTGAATTTTAAATTTGTCAATTAATCATCTATGCTGATTAAGTCTTCATAAACAATTGA	572
Qy	481	ACTTTTGTATTTATAGCCCAATGCTGCAATATTCATATTTAAATTCCTAAAGAAATAAT	540
Db	573	ACTTTTGTATTTATAGCCCAATGCTGCAATATTCATATTTAAATTCCTAAAGAAATAAT	632
Qy	541	TTTAAATGTTAAACGCTGATAATGCAATTAATAGAAAATGTGGTTTACAAAATAAAAACG	600
Db	633	TTTAAATGTTAAACGCTGATAATGCAATTAATAGAAAATGTGGTTTACAAAATAAAAACG	692
Qy	601	GTCTTCACTAGTTTACACCTGAAAGTAAAGTGTCTCGTTTGGAAAGCTAAGAGCCATCAT	660
Db	693	GTCTTCACTAGTTTACACCTGAAAGTAAAGTGTCTCGTTTGGAAAGCTAAGAGCCATCAT	752
Qy	661	GTCTAAGAGTGAACCACTGACAACTGAGAGAGTCAGGACCAACATTTCTGTCTTGAAAAG	720
Db	753	GTCTAAGAGTGAACCACTGACAACTGAGAGAGTCAGGACCAACATTTCTGTCTTGAAAAG	812
Qy	721	AATTGTAACATCATGCTATGGCCCTTCAGTGAAGTGAAGCAGCTGACAAATGGCTTTGG	780
Db	813	AATTGTAACATCATGCTATGGCCCTTCAGTGAAGTGAAGCAGCTGACAAATGGCTTTGG	872
Qy	781	AGGTTACGTGTGTACAACTTCACAGTCTCAGCTCTGCTCAGTCACCTTTTGTGTACACA	840
Db	873	AGGTTACGTGTGTACAACTTCACAGTCTCAGCTCTGCTCAGTCACCTTTTGTGTACACA	932
Qy	841	TCCCATTTTAAAGATCCTGACAGCCTCCATACAGAATCATGTGTCAGGCTTCAGTGATTG	900

Db	933	TCCCATTTTAAAGATCCTGACAGCCTCCATACAGAATCATGTGTCAGGCTTCAGTGATTG	992
Qy	901	TGCTTTATTTACAGCTATTTCTTTTGTGCAACCTGATTGAAAATGTTTCAGAGATTAGGCTT	960
Db	993	TGCTTTATTTACAGCTATTTCTTTTGTGCAACCTGATTGAAAATGTTTCAGAGATTAGGCTT	1052
Qy	961	GACACCCACCATCTGTCAITTAGATTAAATAAACAATCTTTTGGAGCTTTTGGCATCAGTTATCT	1020
Db	1053	GACACCCACCATCTGTCAITTAGATTAAATAAACAATCTTTTGGAGCTTTTGGCATCAGTTATCT	1112
Qy	1021	CAAGTCTGATACCTGTGGTTTGCAGATCCAGTGGACTTTTAGTAGTACTCAGATCTCTCT	1080
Db	1113	CAAGTCTGATACCTGTGGTTTGCAGATCCAGTGGACTTTTAGTAGTACTCAGATCTCTCTCT	1172
Qy	1081	TTGTTTGGTGGTGTAGTATATTAAACAAGTAAACCTGCTGTATGCTCACCAGAAAGGAAC	1140
Db	1173	TTGTTTGGTGGTGTAGTATATTAAACAAGTAAACCTGCTGTATGCTCACCAGAAAGGAAC	1232
Qy	1141	AGAGCATGTGAGTGTCTTGCATCCTGAGAGCCTTTTTCCTTACAAATTCAGAAAATGCTGA	1200
Db	1233	AGAGCATGTGAGTGTCTTGCATCCTGAGAGCCTTTTTCCTTACAAATTCAGAAAATGCTGA	1292
Qy	1201	AGGCCACATCATTTTAAAGAAAGATTTAAATTTGTAACCTTTTAAAGGTCAGAGATTATAGA	1260
Db	1293	AGGCCACATCATTTTAAAGAAAGATTTAAATTTGTAACCTTTTAAAGGTCAGAGATTATAGA	1352
Qy	1261	TTCCACTGTATTACCTGGGATCTCAATGAAATGTGAGAGTTCAAATTAATAGGCTTAT	1320
Db	1353	TTCCACTGTATTACCTGGGATCTCAATGAAATGTGAGAGTTCAAATTAATAGGCTTAT	1412
Qy	1321	ACCTATCAAAAAATCAATGCTCCCTCAAGGTGGGACCTCTTTTGTACAACTTTATCCCGAGA	1380
Db	1413	ACCTATCAAAAAATCAATGCTCCCTCAAGGTGGGACCTCTTTTGTACAACTTTATCCCGAGA	1472
Qy	1381	CACTTCTGACACTGGAGAAAGAACTGTGTGTGTCAGTTATGGGTTTCTCTTCAAAATGC	1440
Db	1473	CACTTCTGACACTGGAGAAAGAACTGTGTGTGTCAGTTATGGGTTTCTCTTCAAAATGC	1532
Qy	1441	AGTCTTGGACACAGCTGCTTAACTAGGAAGGAGCTTAATCAGTGACCACTAGATCTTGT	1500
Db	1533	AGTCTTGGACACAGCTGCTTAACTAGGAAGGAGCTTAATCAGTGACCACTAGATCTTGT	1592
Qy	1501	CCTGTGCCAAAAAGTTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGATCATCGTATTAT	1560
Db	1593	CCTGTGCCAAAAAGTTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGATCATCGTATTAT	1652
Qy	1561	TGCCATAGACAGAAATTTGGAGTGACTCTGTATGGAAACCCCTGACTAAAATGACAGGAAACACA	1620
Db	1653	TGCCATAGACAGAAATTTGGAGTGACTCTGTATGGAAACCCCTGACTAAAATGACAGGAAACACA	1712
Qy	1621	GCCTATTGGATCCCTAGGCTCAATATGCTCCTAATAGTTATGGAGTGTGAAAGATGTGTG	1680
Db	1713	GCCTATTGGATCCCTAGGCTCAATATGCTCCTAATAGTTATGGAGTGTGAAAGATGTGTG	1772
Qy	1681	CACTGCAAAATTTGGCTCCCAACATTTTTCATCTTATTTCTTAATGAAAGCAACAATCTG	1740
Db	1773	CACTGCAAAATTTGGCTCCCAACATTTTTCATCTTATTTCTTAATGAAAGCAACAATCTG	1832
Qy	1741	CAGCTTGTCTTCTGCAACAGAAAATGACACTGCTGGGATGAGCTGAAGCTCACGCTGTCAC	1800
Db	1833	CAGCTTGTCTTCTGCAACAGAAAATGACACTGCTGGGATGAGCTGAAGCTCACGCTGTCAC	1892
Qy	1801	GACGGCAGCTGCTGCTGAGTTAACTCAAGGNAACCATGGGCTTTGTTGGAGGTGG	1860
Db	1893	GACGGCAGCTGCTGCTGAGTTAACTCAAGGNAACCATGGGCTTTGTTGGAGGTGG	1952
Qy	1861	CTGTACTGAAAATCTCAATTTGGCTGCAATATCAGACAAAGACTCACAAACGCCAGAAAG	1920
Db	1953	CTGTACTGAAAATCTCAATTTGGCTGCAATATCAGACAAAGACTCACAAACGCCAGAAAG	2012
Qy	1921	CAATCTCAAAGATGATGATGATCTCAAAACAGAACTTCAATTAATTTGCTGAAGCATTTTG	1980
Db	2013	CAATCTCAAAGATGATGATGATCTCAAAACAGAACTTCAATTAATTTGCTGAAGCATTTTG	2072

QY 1981 CAGTGCCTAGATCTGTGTGGCTCTTTAGAACATGATGAGGTGAAATTCCTACTGA 2040
|||||
Db 2073 CAGTGCCTAGATCTGTGTGGCTCTTTAGAACATGATGAGGTGAAATTCCTACTGA 2132
|||||
QY 2041 CATGAAGTATGGACACCTTTGGTCAGTTCAGGCAGATTCCTCCTGTGTGGTAACTGGCC 2100
|||||
Db 2133 CATGAAGTATGGACACCTTTGGTCAGTTCAGGCAGATTCCTCCTGTGTGGTAACTGGCC 2192
|||||
QY 2101 AGATTTGCTTTACAGTGTGGCTGTGGATTTATACATACCCAGGAAGAACTCAACTGGTC 2160
|||||
Db 2193 AGATTTGCTTTACAGTGTGGCTGTGGATTTATACATACCCAGGAAGAACTCAACTGGTC 2252
|||||
QY 2161 TTTCTTTAAGACAGACGCTGTCATTTGTGGCCAGAAAGTGCCTTCCACATGAAGCTGT 2220
|||||
Db 2253 TTTCTTTAAGACAGACGCTGTCATTTGTGGCCAGAAAGTGCCTTCCACATGAAGCTGT 2312
|||||
QY 2221 GGCTCAGCCAGCAACCTGACCTTGACTGTGGACTTTTGACTGCMAAGCTTAGTGGCTACAGGT 2280
|||||
Db 2313 GGCTCAGCCAGCAACCTGACCTTGACTGTGGACTTTTGACTGCMAAGCTTAGTGGCTACAGGT 2372
|||||
QY 2281 GGCTGTAGACAGCCAAATTTGATTTGGGATCTTTTCATATGTTATTGAAGATAAAAACTA 2340
|||||
Db 2373 GGCTGTAGACAGCCAAATTTGATTTGGGATCTTTTCATATGTTATTGAAGATAAAAACTA 2432
|||||
QY 2341 AGNATAGCATGTTTCGTATTACAGAGAAACAAATAAACTAGTCTGTTGGCAATTGAGA 2400
|||||
Db 2433 AGAATAGCATGTTTCGTATTACAGAGAAACAAATAAACTAGTCTGTTGGCAATTGAAA 2492
|||||
QY 2401 AAAAAAAAAAAAAAAAAAAAA 2420
|||||
Db 2493 GAAAAAAAAAAAAAAAAAAAA 2512
|||||

RESULT 5
AF221992 2516 bp mRNA linear PRI 02-JUN-2000
LOCUS Homo sapiens putative chaperonin MKKS (MKKS) mRNA, complete cds,
DEFINITION alternatively spliced.
ACCESSION AF221992
VERSION AF221992.1 GI:8163831
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2516)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Stone,D.L., Slavotinek,A., Bouffard,G.G., Banerjee-Basu,S.,
Baxevarian,A.D., Barr,M. and Biesecker,L.G.
TITLE Mutation of a gene encoding a putative chaperonin causes
McKusick-Kaufman syndrome
JOURNAL Nat. Genet. 25 (1), 79-82 (2000)
MEDLINE 20264373
PUBMED 10802661
REFERENCE 2 (bases 1 to 2516)
AUTHORS Biesecker,L.G., Slavotinek,A.M. and Stone,D.L.
TITLE Direct submission
JOURNAL Submitted (07-JAN-2000) LGDR, NHGRI/NIH, 49 Convent Drive,
Bethesda, MD 20892-4472, USA
FEATURES
source
1..2516
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="20p12"
/note="corresponds to genomic sequence deposited in
GenBank Accession Number AL034430"
1..2516
/gene="MKKS"
/notes="McKusick-Kaufman (hydrometrocopolos-polydactyly)
syndrome"
736..2448

/gene="MKKS"
/note="transcript MKKSa; includes alternative exon 1A"
/codon_start=1
/product="putative chaperonin MKKS"
/protein_id="AAF73872.1"
/db_xref="GI:8163832"
/translation="MSRLEAKKPSLCKSEPLTTERVRTTSLVKRIVTSYGPSRLK
QLHNGFGYVCTTSOSSALLSHLLVTHPIKLITFASIQNHVSFSDGGLFTAILCCNL
IENVRGLGTPVIRLNKHLISLCISYKSETCGCRIPVDFSSQIILLCULSVILTS
KPAKMLTKRETHVSALILRAFLPIPENAEHIIILGSLIVPLKQGVSDVSLVPGI
LTIEMSEVOLMRLLPIKSTALKVALFCTTSGDSTGEGTVVVGYSLENAVLDQI
LNLGRLISDHDVLVCKQIHPSLKQFLNHRITAIIDIGVTLMEPLTKMTGTQPIG
SIGSTCPNSYGVKDVCTAKFGSKHFFHLIPNEATICSLILCNRNDTAWDELKTCOT
ALHVLQILFLKEPALLGGCTETHLAAVIRKHTNDPESILKDDSCCTQELQIIAEAF
CSALRSVVGSLHEDGGEILTDKYLHLSVQADSPCVANWPDLLSOCGGLNSQEEL
NWSFLRSTRPFPVQSCLPHEAVGSASNLTLDCILTAKLSGLQAVAVETANLILDLSTVI
EDKN"

ORIGIN

Query Match	97.1%;	Score 2401;	DB 9;	Length 2516;
Best Local Similarity	99.8%;	Pred. No. 0;		
Matches 2404;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;

QY 1 TATGAGCCCTTCGGAACCTTGTGAGAGACTACAAAGTTTGTGGTTGTATGGTCCCTTAGT 60
|||||
Db 108 TATGAGCCCTTCGGAACCTTGTGAGAGACTACAAAGTTTGTGGTTGTATGGTCCCTTAGT 167
|||||
QY 61 TGGGCTCATACATTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 120
|||||
Db 168 TGGGCTCATACATTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 227
|||||
QY 121 AAACGACGACATTCCTGACCAAGATAGTCTGGGACTTTCAAATCTTCAGAAAGCCAAAT 180
|||||
Db 228 AAACGACGACATTCCTGACCAAGATAGTCTGGGACTTTCAAATCTTCAGAAAGCCAAAT 287
|||||
QY 181 CCAGGGAGTAGCAGGCTTGCAATCTTCAGGTAAAGAGCAGCTTTTGAATCTGAGCTTC 240
|||||
Db 288 CCAGGGAGTAGCAGGCTTGCAATCTTCAGGTAAAGAGCAGCTTTTGAATCTGAGCTTC 347
|||||
QY 241 ATATCGAAAGAGAGATGAAAAATACCAAGTTGGAATAGAAAGAACTGGGCTTCTTGAGCT 300
|||||
Db 348 ATATCGAAAGAGAGATGAAAAATACCAAGTTGGAATAGAAAGAACTGGGCTTCTTGAGCT 407
|||||
QY 301 GGGATATCTTTTCATAGGTGCCATCTTGGAAACATATCTTTTCGAGAGGCTCCGAAGCAG 360
|||||
Db 408 GGGATATCTTTTCATAGGTGCCATCTTGGAAACATATCTTTTCGAGAGGCTCCGAAGCAG 467
|||||
QY 361 TCTGTAAAAATTTTCAGTCTCAAGCAACAAAGAGATTTTGAAGAGTGAAGTAAAAATAAT 420
|||||
Db 468 TCTGTAAAAATTTTCAGTCTCAAGCAACAAAGAGATTTTGAAGAGTGAAGTAAAAATAAT 527
|||||
QY 421 ATTTGGAATTAATAATTTGTTCATTAATCATTTCTATGCTGATTAAGCTTCATAAACATTGA 480
|||||
Db 528 ATTTGGAATTAATAATTTGTTCATTAATCATTTCTATGCTGATTAAGCTTCATAAACATTGA 587
|||||
QY 481 ACTTTTGTATTTTATAGCCCAATGCTGCATATTCATATCTTAATTCCTAAAGAAATAAT 540
|||||
Db 588 ACTTTTGTATTTTATAGCCCAATGCTGCATATTCATATCTTAATTCCTAAAGAAATAAT 647
|||||
QY 541 TTTTAATGTTAAAAACGTGATAATGCAATAATAAGAAAAATGTGGTTTACAAAAATAAAACG 600
|||||
Db 648 TTTTAATGTTAAAAACGTGATAATGCAATAATAAGAAAAATGTGGTTTACAAAAATAAAACG 707
|||||
QY 601 GTCTTCACCTAGTTTACCACTGAGTAAGATGTCTCGTTTGAAGCTGAAGAGCCATCAT 660
|||||
Db 708 GTCTTCACCTAGTTTACCACTGAGTAAGATGTCTCGTTTGAAGCTGAAGAGCCATCAT 767
|||||
QY 661 GTGTAAGAGTGAACCACTGACAACTGAGAGAGTCTAGGACCACTTCTGCTCTTGAAGAG 720
|||||
Db 768 GTGTAAGAGTGAACCACTGACAACTGAGAGAGTCTAGGACCACTTCTGCTCTTGAAGAG 827
|||||
QY 721 AATTGTAAACATCATGCTATGGCCCTTCAGGTAGGCTGAAGAGAGCTGCACAAATGGCTTTGG 780
|||||

Db 828 AATTGTTAAATCATCTGCTTANGSCCCCTCAGGTAGGCTGAAGCAGCTGCACAATGGCTTTGG 887
Qy 781 AGGTTACGCTGTGTAAACCTCACAGTCTCTCAGCTCTGCTCAGTCAACCTTTTGGTTCACACA 840
Db 888 AGGTTACGCTGTGTAAACCTCACAGTCTCTCAGCTCTGCTCAGTCAACCTTTTGGTTCACACA 947
Qy 841 TCCCAATTTTAAAGATCTCTGACAGCTCTCATACAGAAATCATGTGTCAAGCTTCTAGTGATTG 900
Db 948 TCCCAATTTTAAAGATCTCTGACAGCTCTCATACAGAAATCATGTGTCAAGCTTCTAGTGATTG 1007
Qy 901 TGGCTTATTCACAGCTATCTTTTCTGCAACCTGATTTGAAATGTTTCTCAGAGATTAGGCTT 960
Db 1008 TGGCTTATTCACAGCTATCTTTTCTGCAACCTGATTTGAAATGTTTCTCAGAGATTAGGCTT 1067
Qy 961 GACACCCACCACTGCTCATTTAGATTAATAAATCAATCTTTTGGTCTTTTGGCATCAGTTATCT 1020
Db 1068 GACACCCACCACTGCTCATTTAGATTAATAAATCAATCTTTTGGTCTTTTGGCATCAGTTATCT 1127
Qy 1021 CAAGTCTGATACCTGTGGTGTGCAATCCAGTGGACTTTTAGTAGTACTCAGATCTCTCT 1080
Db 1128 CAAGTCTGAGACCTGTGGTGTGCAATCCAGTGGACTTTTAGTAGTACTCAGATCTCTCT 1187
Qy 1081 TTGTTTCTGCTGCTAGTATATTAACAAGTAAACCTGCTGTATGCTCACAGAAAGGAAC 1140
Db 1188 TTGTTTCTGCTGCTAGTATATTAACAAGTAAACCTGCTGTATGCTCACAGAAAGGAAC 1247
Qy 1141 AGAGCATGTCTGATCTCTGAGAGCTCTTTTGTCTTACAATTCAGAAATGCTGA 1200
Db 1248 AGAGCATGTCTGATCTCTGAGAGCTCTTTTGTCTTACAATTCAGAAATGCTGA 1307
Qy 1201 AGGCCACATCAATTTTAGGAAAGAGTTAAATGTACCTTTTAAAGGTCAGAAAGTTATAGA 1260
Db 1308 AGGCCACATCAATTTTAGGAAAGAGTTAAATGTACCTTTTAAAGGTCAGAAAGTTATAGA 1367
Qy 1261 TTCACATGATTAACCTGGGATCTCAATGAAATGTCAGAAATCAATTAATAGGCTATT 1320
Db 1368 TTCACATGATTAACCTGGGATCTCAATGAAATGTCAGAAATCAATTAATAGGCTATT 1427
Qy 1321 ACCTATCAAAAAATCAATGCTCCCTCAAGGTGGCACTCTTTGTACAACTTTATCCGGAGA 1380
Db 1428 ACCTATCAAAAAATCAATGCTCCCTCAAGGTGGCACTCTTTGTACAACTTTATCCGGAGA 1487
Qy 1381 CACTTCTGACATGGAAGAAATCTGTGGTGGTCAAGTTATGGGGTTCTCTTCAAAATGC 1440
Db 1488 CACTTCTGACATGGAAGAAATCTGTGGTGGTCAAGTTATGGGGTTCTCTTCAAAATGC 1547
Qy 1441 AGTCTTGGACCACTGCTTAACTTAGAAGGCACTTAATCAGTGACCACTAGATCTTGT 1500
Db 1548 AGTCTTGGACCACTGCTTAACTTAGAAGGCACTTAATCAGTGACCACTAGATCTTGT 1607
Qy 1501 CCTGTGCAAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGATCATCTATTAT 1560
Db 1608 CCTGTGCAAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGATCATCTATTAT 1667
Qy 1561 TGGCATAGACAGAAATTTGAGTGTACTCTGATGGAACCCCTGACTTAAATAGCAGGAACACA 1620
Db 1668 TGGCATAGACAGAAATTTGAGTGTACTCTGATGGAACCCCTGACTTAAATAGCAGGAACACA 1727
Qy 1621 GCCTATTGGATCCCTAGGCTCAATATGCTCTTAATAGTTATGGAAGTGTGAAGATGTGTG 1680
Db 1728 GCCTATTGGATCCCTAGGCTCAATATGCTCTTAATAGTTATGGAAGTGTGAAGATGTGTG 1787
Qy 1681 CACTGCAAAATTTGGCTCCCAACATTTTTTTTCACTTTATCTTAATGAAGCAACATCTG 1740
Db 1788 CACTGCAAAATTTGGCTCCCAACATTTTTTTTCACTTTATCTTAATGAAGCAACATCTG 1847
Qy 1741 CAGCTTGCTTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCACGTGTCA 1800
Db 1848 CAGCTTGCTTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCACGTGTCA 1907
Qy 1801 GAGGGCACTGATCTCTGCAAGTTAACTCAAGGAACCAATGGCTTTTGGGAGGTGG 1860
Db 1908 GAGGGCACTGATCTCTGCAAGTTAACTCAAGGAACCAATGGCTTTTGGGAGGTGG 1967

Qy 1861 CTGTACTGAAACTCATTTTGGCTCATATATCAGACACAGACTCACAAACGACCCAGAAAG 1920
Db 1968 CTGTACTGAAACTCATTTTGGCTCATATATCAGACACAGACTCACAAACGACCCAGAAAG 2027
Qy 1921 CATTTCTCAAAGATGATGAATGTACTCAACAGAACTTCAATTAATTTGCTGAAGCATTTTG 1980
Db 2028 CATTTCTCAAAGATGATGAATGTACTCAACAGAACTTCAATTAATTTGCTGAAGCATTTTG 2087
Qy 1981 CAGTGGCCCTAGAAATCTGTTTGGCTCTTTAGAACATGATGAGGTTGAAATTTCTCACTGA 2040
Db 2088 CAGTGGCCCTAGAAATCTGTTTGGCTCTTTAGAACATGATGAGGTTGAAATTTCTCACTGA 2147
Qy 2041 CATGAAAGTATGGACACCTTTTGGTCTGAGTTCAGGAGGATTTCTCCCTGCTGTCTAACTGGCC 2100
Db 2148 CATGAAAGTATGGACACCTTTTGGTCTGAGTTCAGGAGGATTTCTCCCTGCTGTCTAACTGGCC 2207
Qy 2101 AGATTTGCTTTTACAGTGTGGCTGTGGATTTATACAATAGCAGGAAGAACTCAACTGGTC 2160
Db 2208 AGATTTGCTTTTACAGTGTGGCTGTGGATTTATACAATAGCAGGAAGAACTCAACTGGTC 2267
Qy 2161 TTTCTTAAAGAACACACGCTGCTCCATTTTGTGGCCACAAAGCTGCTTCCCATGAGCTGT 2220
Db 2268 TTTCTTAAAGAACACACGCTGCTCCATTTTGTGGCCACAAAGCTGCTTCCCATGAGCTGT 2327
Qy 2221 GGGCTCAGCCAGCAACCTTGACCTTGGACTGTTTGAAGCTTGTAGTGGCTTACAGGT 2280
Db 2328 GGGCTCAGCCAGCAACCTTGACCTTGGACTGTTTGAAGCTTGTAGTGGCTTACAGGT 2387
Qy 2281 GGGCTCAGCCAGCAACCTTTGATTTGGGATCTTTTCAATGTTTATTTGAAGATAAAACTA 2340
Db 2388 GGGCTCAGCCAGCAACCTTTGATTTGGGATCTTTTCAATGTTTATTTGAAGATAAAACTA 2447
Qy 2341 AGAGATAGCATGTTTGGTATTACAGAGAAACAAATAAAGTCTGTTGGCAATTCAGA 2400
Db 2448 AGAGATAGCATGTTTGGTATTACAGAGAAACAAATAAAGTCTGTTGGCAATTCAGA 2507
Qy 2401 AAAAAAAAAA 2409
Db 2508 AAAAAAAAAA 2516

RESULT 6
AK025741
LOCUS Homo sapiens cDNA: FLJ22088 fis, clone HEP15963. 2503 bp mRNA linear PRI 13-SEP-2003
DEFINITION AK025741
ACCESSION AK025741
VERSION AK025741.1 GI:10438352
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2503)
AUTHORS Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T., Shibahara, T., Tanaka, T. and Nakamura, Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail: fldcna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science,

University of Tokyo (partly supported by Science and Technology Agency).

FEATURES	source	Location/Qualifiers
CDS	1..2503	/organism="Homo sapiens"
		/molecule="mRNA"
		/db_xref="taxon:9606"
		/clone="HEP15963"
		/cell_line="HepG2"
		/cell_type="hepatoma"
		/clone_lib="HEP"
		/notes="cloning vector pME18SFL3"
		716..2428
		/notes="unnamed protein product"
ORIGIN	Query Match	97.0%; Score 2398.4; DB 9; Length 2503;
		Best Local Similarity 99.5%; Pred. No. 0;
		Matches 2405; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
		1 TATGAGCCTTCGGAACCTGTGGAGAGACTACAAAGTTTGGTGTGTTATGTCCTCTTAGT 60
		88 TATGAGCCTTCGGAACCTGTGGAGAGACTACAAAGTTTGGTGTGTTATGTCCTCTTAGT 147
		61 TGGGCTCATACATTTGGGGTGTGACAGATCAAAAGCAGCCCTGTTTCCAAATACCTAA 120
		148 TGGGCTCATACATTTGGGGTGTGACAGATCAAAAGCAGCCCTGTTTCCAAATACCTAA 207
		121 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACATTTCAAATCTTCAGAAGAGCCAAAT 180
		208 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACATTTCAAATCTTCAGAAGAGCCAAAT 267
		181 CCAGGGGAAGTAGCAGCTTGCATCTTCAGGTAAAGAGCAGCTTTGAAATCTGAGCTTC 240
ORIGIN	Query Match	97.0%; Score 2398.4; DB 9; Length 2503;
		Best Local Similarity 99.5%; Pred. No. 0;
		Matches 2405; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
		1 TATGAGCCTTCGGAACCTGTGGAGAGACTACAAAGTTTGGTGTGTTATGTCCTCTTAGT 60
		88 TATGAGCCTTCGGAACCTGTGGAGAGACTACAAAGTTTGGTGTGTTATGTCCTCTTAGT 147
		61 TGGGCTCATACATTTGGGGTGTGACAGATCAAAAGCAGCCCTGTTTCCAAATACCTAA 120
		148 TGGGCTCATACATTTGGGGTGTGACAGATCAAAAGCAGCCCTGTTTCCAAATACCTAA 207
		121 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACATTTCAAATCTTCAGAAGAGCCAAAT 180
		208 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACATTTCAAATCTTCAGAAGAGCCAAAT 267
		181 CCAGGGGAAGTAGCAGCTTGCATCTTCAGGTAAAGAGCAGCTTTGAAATCTGAGCTTC 240

QY	601	GTCTTCACTAGTTACCACTGAAGTAAGTGTCTGTTTGGAGCTTAAGAGCCATCATTT 660
DB	688	GTCTTCACTAGTTACCACTGAAGTAAGTGTCTGTTTGGAGCTTAAGAGCCATCATTT 747
QY	661	GTGTAAGAGTGAACCACTGACAACTGAGAGAGTCAAGGACCACTTTCTGCTTTGAAAAG 720
DB	748	GTGTAAGAGTGAACCACTGACAACTGAGAGAGTCAAGGACCACTTTCTGCTTTGAAAAG 807
QY	721	AATTGTAAACATCATGCTATGGCCCTCAGGTAGGTGAAGCAGCTGCACAATGGCTTTGG 780
DB	808	AATTGTAAACATCATGCTATGGCCCTCAGGTAGGTGAAGCAGCTGCACAATGGCTTTGG 867
QY	781	AGGTACAGTGTGTACAACTCAGTCTCAGCTCTGCTCAGTCACTTTTGGTCAACA 840
DB	868	AGGTACAGTGTGTACAACTCAGTCTCAGCTCTGCTCAGTCACTTTTGGTCAACA 927
QY	841	TCCCATTTTAAAGATCCTGACAGCCTCCATACAGAAATCATGTGTCAAGCTTCAGTGATTG 900
DB	928	TCCCATTTTAAAGATCCTGACAGCCTCCATACAGAAATCATGTGTCAAGCTTCAGTGATTG 987
QY	901	TGGCTTATTTACAGCTATTTCTTTTGTGCAACCTGATTGAAAATGTTTCAGAGATTAGGCTT 960
DB	988	TGGCTTATTTACAGCTATTTCTTTTGTGCAACCTGATTGAAAATGTTTCAGAGATTAGGCTT 1047
QY	961	GACACCCACCACTGTGATTAATTAATAAATCATCTTTTGTGCTTTTGGTCACTTGTATCT 1020
DB	1048	GACACCCACCACTGTGATTAATTAATAAATCATCTTTTGTGCTTTTGGTCACTTGTATCT 1107
QY	1021	CAAGTCTGATACCTGTTGTCGAATCCAGTGGACTTTTGTAGTACTCAGATCTCTCTCT 1080
DB	1108	CAAGTCTGATACCTGTTGTCGAATCCAGTGGACTTTTGTAGTACTCAGATCTCTCTCT 1167
QY	1081	TTGTTTGGTGTGATTAATTAATAAATCATCTTTTGTGCTTTTGGTCACTTGTATCT 1140
DB	1168	TTGTTTGGTGTGATTAATTAATAAATCATCTTTTGTGCTTTTGGTCACTTGTATCT 1227
QY	1141	AGAGCATGTGATGCTTTGATCTGAGAGCCTTTTGTGCTTCAATTCAGAAATGCTCA 1200
DB	1228	AGAGCATGTGATGCTTTGATCTGAGAGCCTTTTGTGCTTCAATTCAGAAATGCTCA 1287
QY	1201	AGGCGACATCATTTTGAAGAGATTTAATTTGCTTACCTTTTAAAGGTCAGAGATTATAGA 1260
DB	1288	AGGCGACATCATTTTGAAGAGATTTAATTTGCTTACCTTTTAAAGGTCAGAGATTATAGA 1347
QY	1261	TTCCACTGTATTTACCTGGGATCTCAATTTGAAGATGTCAGAGTTCAATTAATGAGGCTATT 1320
DB	1348	TTCCACTGTATTTACCTGGGATCTCAATTTGAAGATGTCAGAGTTCAATTAATGAGGCTATT 1407
QY	1321	ACCTATCAAAAATCAATGCTCCCTCAAGTGGCACTCTTTTGTACAACTTTATCCGGAGA 1380
DB	1408	ACCTATCAAAAATCAATGCTCCCTCAAGTGGCACTCTTTTGTACAACTTTATCCGGAGA 1467
QY	1381	CACTTCTGACACTGGAGAGAGAACTGTGGTGGTCAAGTATGGGTTTCTCTTGAATAATGC 1440
DB	1468	CACTTCTGACACTGGAGAGAGAACTGTGGTGGTCAAGTATGGGTTTCTCTTGAATAATGC 1527
QY	1441	AGTCTTGGACCAAGTCTTAACTTAGGAAGGAGCTTAATCAGTGACCACTGATCTTGT 1500
DB	1528	AGTCTTGGACCAAGTCTTAACTTAGGAAGGAGCTTAATCAGTGACCACTGATCTTGT 1587
QY	1501	CCTGTGCCAAAAGTTTATACATCCATCTTTTGAAGCAGTTTCTCAATATCATCTGATTTAT 1560
DB	1588	CCTGTGCCAAAAGTTTATACATCCATCTTTTGAAGCAGTTTCTCAATATCATCTGATTTAT 1647
QY	1561	TGCCATAGACAGAAATGGAGTGACTCTGATGGAAACCCCTGACTTAAATGACAGGAACACA 1620
DB	1648	TGCCATAGACAGAAATGGAGTGACTCTGATGGAAACCCCTGACTTAAATGACAGGAACACA 1707
QY	1621	GCTTATTTGATCCCTAGGCTCAATATGCTTAAATAGTTTGAAGAGTGTGAAGATGTGTG 1680
DB	1708	GCTTATTTGATCCCTAGGCTCAATATGCTTAAATAGTTTGAAGAGTGTGAAGATGTGTG 1767

Db 786 GTGTAAGAGTGAACTGACAACTGAGAGAGTCAGGACCAACACATTTCTGTCTTGAAGAG 845

Qy 721 AATTGTAAACATCATGTATGCGCCCTCAGTAGGCTGAAGAGAGCTGCACAATGGCTTTGG 780

Db 846 AATTGTAAACATCATGTATGCGCCCTCAGTAGGCTGAAGAGAGCTGCACAATGGCTTTGG 905

Qy 781 AGTTACGCTGTGACAACTCAGAGTCCCTCAGCTCTGCTCAGTCACTTTTGGTACACA 840

Db 906 AGTTACGCTGTGACAACTCAGAGTCCCTCAGCTCTGCTCAGTCACTTTTGGTACACA 965

Qy 841 TCCATTTTAAAGATCCTGACAGCTCCATACAGAAATCATGTGTCAAGCTTCAGTGATTG 900

Db 966 TCCATTTTAAAGATCCTGACAGCTCCATACAGAAATCATGTGTCAAGCTTCAGTGATTG 1025

Qy 901 TGGCTTATTACAGCTATTCTTTGCTGCACTGATTGAAATGTTACAGATTAGGCTT 960

Db 1026 TGGCTTATTACAGCTATTCTTTGCTGCACTGATTGAAATGTTACAGATTAGGCTT 1085

Qy 961 GACACCCACACCTGCTATTAGATTAAATAAACATCTTTTGAGTCTTTTGATCATGTTATCT 1020

Db 1086 GACACCCACACCTGCTATTAGATTAAATAAACATCTTTTGAGTCTTTTGATCATGTTATCT 1145

Qy 1021 CAACTCTGATACCTGTGGTGTGCGAATCCAGTGGACTTTTAGTAGTACTCAGATCCTCT 1080

Db 1146 CAACTCTGATACCTGTGGTGTGCGAATCCAGTGGACTTTTAGTAGTACTCAGATCCTCT 1205

Qy 1081 TTGTTTGGTGTAGTATATTAAAGATTAACCTGCTGTATGCTCACCAGAAAGAAAC 1140

Db 1206 TTGTTTGGTGTAGTATATTAAAGATTAACCTGCTGTATGCTCACCAGAAAGAAAC 1265

Qy 1141 AGAGCATGTCACTGCTGCTGAGAGCTTTTGGCTTTTACAAATCCAGAAATGCTCA 1200

Db 1266 AGAGCATGTCACTGCTGAGAGCTTTTGGCTTTTACAAATCCAGAAATGCTCA 1325

Qy 1201 AGGCCACATCATTTTAGGAAAGAGTTTAAATGTGACCTTTTAAAGGTCAAGAGTTATAGA 1260

Db 1326 AGGCCACATCATTTTAGGAAAGAGTTTAAATGTGACCTTTTAAAGGTCAAGAGTTATAGA 1385

Qy 1261 TTCCACTGTATTACCTGGGATCTCAATGAAATGTGAGAGTTCAATTAATGAGGCTATT 1320

Db 1386 TTCCACTGTATTACCTGGGATCTCAATGAAATGTGAGAGTTCAATTAATGAGGCTATT 1445

Qy 1321 ACCTATCAAAAATCAACTGCGCTCAAGGTGGGACCTTTTGTACAACTTTTATCCGGAGA 1380

Db 1446 ACCTATCAAAAATCAACTGCGCTCAAGGTGGGACCTTTTGTACAACTTTTATCCGGAGA 1505

Qy 1381 CACTTCTGACACTGGAGAGGAACTGTGTGTGTGCTGATGAGGTTTCTTTGAAAATGC 1440

Db 1506 CACTTCTGACACTGGAGAGGAACTGTGTGTGTGCTGATGAGGTTTCTTTGAAAATGC 1565

Qy 1441 AGTCTTGGACCACTGCTTAACTAGGAAGGAGCTTAATCAGTGACCACTGATCTTGT 1500

Db 1566 AGTCTTGGACCACTGCTTAACTAGGAAGGAGCTTAATCAGTGACCACTGATCTTGT 1625

Qy 1501 CTTGTGCCAAAAGTTATACATCCATCTTTGAAGCAGTTTCTCAATGATCATCGTATTAT 1560

Db 1626 CTTGTGCCAAAAGTTATACATCCATCTTTGAAGCAGTTTCTCAATGATCATCGTATTAT 1685

Qy 1561 TGCATATAGACAGAAATTTGGAGTACTCTGATGGAAACCCCTGACTTAAATGACAGGAA 1620

Db 1686 TGCATATAGACAGAAATTTGGAGTACTCTGATGGAAACCCCTGACTTAAATGACAGGAA 1745

Qy 1621 GCCTATTGATCCCTAGGCTCAATATGTCCTAATAGTTATGGAAGTGTGAAGATGTGTG 1680

Db 1746 GCCTATTGATCCCTAGGCTCAATATGTCCTAATAGTTATGGAAGTGTGAAGATGTGTG 1805

Qy 1681 CACTGCAAAATTTGGCTCCCAACATTTTTCATCTTATTCTTAAATGAAGCAACAATCTG 1740

Db 1806 CACTGCAAAATTTGGCTCCCAACATTTTTCATCTTAAATGAAGCAACAATCTG 1865

Qy 1741 CAGCTTGTCTCTGCAACAGAAATGACACTG 1772

Db 1866 CAGCTTGTCTCTGCAACAGAAATGAGGGTG 1897

RESULT 8
CQ729071
LOCUS 1713 bp DNA linear PAT 03-FEB-2004
Sequence 15005 from Patent WO02068579.
CQ729071
ACCESSION
CQ729071.1 GI:42299608
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL
Patent: WO 02068579-A 15005 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source
location/Qualifiers
1..1713
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 69.1%; Score 1708.2; DB 6; Length 1713;
Best Local Similarity 99.8%; Pred. No. 7.3e-289;
Matches 1710; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 629 ATGTCTCGTTTGGAGCTTAAGAGCCATCATTTGTGTAGAGTGAAACCCTGACCACTGAG 688
Db 1 ATGTCTCGTTTGGAGCTTAAGAGCCATCATTTGTGTAGAGTGAAACCCTGACCACTGAG 60
Qy 689 AGAGTCAGGACCAACACATTTCTCTCTTTGAAAGAAATTTGTAACATCATGCTATGGCCCTCA 748
Db 61 AGAGTCAGGACCAACACATTTCTCTCTTTGAAAGAAATTTGTAACATCATGCTATGGCCCTCA 120
Qy 749 GGTAGCTGAAGCAGCTGCACAATGGCTTTGGAGTTACGTGTGTACAACTTCACAGTCC 808
Db 121 GGTAGCTGAAGCAGCTGCACAATGGCTTTGGAGTTACGTGTGTACAACTTCACAGTCC 180
Qy 809 TCAGCTCTGCTCAGTCACCTTTTGGTGCACATCCATTTTAAAGATCTGACAGCCTCC 868
Db 181 TCAGCTCTGCTCAGTCACCTTTTGGTGCACATCCATTTTAAAGATCTGACAGCCTCC 240
Qy 869 ATACAGAAATCATGTGTCAAGCTTCAGTGATTTGGCTTATTCACAGCTATTCTTTGCTGC 928
Db 241 ATACAGAAATCATGTGTCAAGCTTCAGTGATTTGGCTTATTCACAGCTATTCTTTGCTGC 300
Qy 929 AACCTGATTGAAATTTGTCAGAGATTAGGCTTGACACCACCATGTGTCAATTAGATTAAAT 988
Db 301 AACCTGATTGAAATTTGTCAGAGATTAGGCTTTGACACCACCATGTGTCAATTAGATTAAAT 360
Qy 989 AACATCTTTTTCAGCTTTTGGCATCAGTTATCTCAAGTCTGATACCTGCTGTCGAATC 1048
Db 361 AACATCTTTTTCAGCTTTTGGCATCAGTTATCTCAAGTCTGATACCTGCTGTCGAATC 420
Qy 1049 CCAGTGGACTTTTAGTAGTACTCAGATCCCTCTTTGTTGGTGGTAGTAGTATTAACAAGT 1108
Db 421 CCAGTGGACTTTTAGTAGTACTCAGATCCCTCTTTGTTGGTGGTAGTAGTATTAACAAGT 480
Qy 1109 AAACCTGCTGTATGCTCACAGAAAGGAAACAGAGCATGTCAAGTCTTTGATTCCTGAGA 1168
Db 481 AAACCTGCTGTATGCTCACAGAAAGGAAACAGAGCATGTCAAGTCTTTGATTCCTGAGA 540
Qy 1169 GCCTTTTTCGCTTACAAATTCAGAAATGCTGAAGCCACATCATTTTACGAAAGAGTTTA 1228
Db 541 GCCTTTTTCGCTTACAAATTCAGAAATGCTGAAGCCACATCATTTTACGAAAGAGTTTA 600
Qy 1229 ATTTGACCTTTTAAAGGTCAAAGAGTTATAGATTCCATGTATTACCTGGGATCTCATTT 1288

Db	601	ATTGTACCTTTAAAGGTCAAAGAGTTATAGATTCCACTGTATTCCTGGGATFACCTATT	660
Qy	1289	GAATGTGAGAGTCTCAATTAATGAGGCTATTACCTATCAAAAAATCAACTGCGCCCTCAAG	1348
Db	661	GAATGTGAGAGTCTCAATTAATGAGGCTATTACCTATCAAAAAATCAACTGCGCCCTCAAG	720
Qy	1349	GTGGCACTCTTTTGTACAACTTTATCGGAGACACTTCTGACACTCGAGAAGAACTGTG	1408
Db	721	GTGGCACTCTTTTGTACAACTTTATCGGAGACACTTCTGACACTCGAGAAGAACTGTG	780
Qy	1409	GTGGTCAGTTATGGGGTTCTCTTGAAATGAGTCTTGACACGAGTCTTAACTCCTAGGA	1468
Db	781	GTGGTCAGTTATGGGGTTCTCTTGAAATGAGTCTTGACACGAGTCTTAACTCCTAGGA	840
Qy	1469	AGGCAGCTAATCAGTGACACAGTAGATCTTGCTGTGCCAAAAGTTATACATCCATCT	1528
Db	841	AGGCAGCTAATCAGTGACACAGTAGATCTTGCTGTGCCAAAAGTTATACATCCATCT	900
Qy	1529	TTGAAGCAGTTTCTCAATATGCAATCGTATATTGTCATAGACAGAAATGGAGTGACTCTG	1588
Db	901	TTGAAGCAGTTTCTCAATATGCAATCGTATATTGTCATAGACAGAAATGGAGTGACTCTG	960
Qy	1589	ATGGAACCCCTGACTAAATGACAGGACACAGCCTATTGGATCCCTAGGCTCAATATGT	1648
Db	961	ATGGAACCCCTGACTAAATGACAGGACACAGCCTATTGGATCCCTAGGCTCAATATGT	1020
Qy	1649	CCTAATAGTTATGGAAGTGTGAAGATGTGTGCACTGCAAAATTTGGCTCCCAACATTTT	1708
Db	1021	CCTAATAGTTATGGAAGTGTGAAGATGTGTGCACTGCAAAATTTGGCTCCCAACATTTT	1080
Qy	1709	TTTTCATCTTATTCCTAATGAAGCAACAATCTGCAGCTTGCTCTCTGCAACGAAATGAC	1768
Db	1081	TTTTCATCTTATTCCTAATGAAGCAACAATCTGCAGCTTGCTCTCTGCAACGAAATGAC	1140
Qy	1769	ACTGCTGGGATGAGCTGAAGCTCAGTGTGACAGCGCACTGCATGCTCTGCGATTAAACA	1828
Db	1141	ACTGCTGGGATGAGCTGAAGCTCAGTGTGACAGCGCACTGCATGCTCTGCGATTAAACA	1200
Qy	1829	CTCAAGAACCATGGGCTTTGTTGGAGTGGCTGTACTGAACTCATTTGGCTGCATAT	1888
Db	1201	CTCAAGAACCATGGGCTTTGTTGGAGTGGCTGTACTGAACTCATTTGGCTGCATAT	1260
Qy	1889	ATCAGACACAAGACTCACAACGCCAGAAAGCAATTTCTCAAGATGATGAATGTACTCAA	1948
Db	1261	ATCAGACACAAGACTCACAACGCCAGAAAGCAATTTCTCAAGATGATGAATGTACTCAA	1320
Qy	1949	ACGAACCTTCAATTAATGCTGAAGCAATTTGAGTGGCTTAGAATCTGTTGGCTCT	2008
Db	1321	ACGAACCTTCAATTAATGCTGAAGCAATTTGAGTGGCTTAGAATCTGTTGGCTCT	1380
Qy	2009	TTAGAACATGATGGAGTGAATTTCTCACTGACATGAAGTATGACACCTTTGGTCAGTT	2068
Db	1381	TTAGAACATGATGGAGTGAATTTCTCACTGACATGAAGTATGACACCTTTGGTCAGTT	1440
Qy	2069	CAGGCAGATTCTCCCTGTGTGCTAACTGCGCCAGATTTGCTTTCAAGTGTGCTGTGA	2128
Db	1441	CAGGCAGATTCTCCCTGTGTGCTAACTGCGCCAGATTTGCTTTCAAGTGTGCTGTGA	1500
Qy	2129	TTATACAATAGCAGAGAACTCAACTGTGCTTTTCTTAAGAGACACAGTGTGCATTT	2188
Db	1501	TTATACAATAGCAGAGAACTCAACTGTGCTTTTCTTAAGAGACACAGTGTGCATTT	1560
Qy	2189	GTGCCACAAGCTGCTCCACATGAAGTGTGGGCTCAGCCAGCAACCTGACCTTGGAC	2248
Db	1561	GTGCCACAAGCTGCTCCACATGAAGTGTGGGCTCAGCCAGCAACCTGACCTTGGAC	1620
Qy	2249	TGTTTGACTGCAAGCTTAGTGGCTTACAGGTGGCTGTAGACAGCCAAATTTGATTTGG	2308
Db	1621	TGTTTGACTGCAAGCTTAGTGGCTTACAGGTGGCTGTAGACAGCCAAATTTGATTTGG	1680
Qy	2309	GATCTTTTCATATGTTATGAAGATAAAAACTAA	2341
Db	1681	GATCTTTTCATATGTTATGAAGATAAAAACTAA	1713

RESULT 9

BC083863

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BC083863

Rattus norvegicus cDNA clone MGC:95109 IMAGE:7127425, complete cds.

BC083863.1

GI:54035334

MGC.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE

1 (bases 1 to 2972)

STRAUSBERG R.L., FEINGOLD E.A., GROUSE L.H., DERGE J.G., KLAUSNER R.D., COLLINS F., WAGNER L., SCHENKEN C.M., SCHULER G.D., ALTSCHUL S.F., ZEEBERG B., BUETOW K.H., SCHAEFER C.F., BHAT N.K., HOPKINS R.F., JORDAN H., MOORE T., MAX S.I., WANG J., HSIEH F., DIATCHENKO L., MARUSINA K., FARMER A.A., RUBIN G.M., HONG L., STAPLETON M., SOARES M.B., BONALDO M.F., CAAVANT T.L., SCHEETZ T.E., BROWNSTEIN M.J., USLIN T.B., TOSHIYUKI S., CARINCI P., FRANGE C., RAHA S., LOQUELLANO N.A., PETERS G.J., ABRAMSON R.D., MULLAHY S.J., BOSAK S.A., MCEWAN P.J., MCKERNAN K.J., MALEK J.A., GUNARATNE P.H., RICHARDS S., WORLEY K.C., HALE S., GARCIA A.M., GAY L.J., HULYK S.W., VILLALON D.K., MUZNY D.M., SODERGREN E.J., LU X., GIBBS R.A., PAHEY J., HELTON E., KETTENMAN M., MADAN A., RODRIGUES S., SANCHEZ A., WHITING M., MADAN A., YOUNG A.C., SHEVCHENKO Y., BOUFFARD G.G., BLAKESLEY R.W., TOUCHMAN J.W., GREEN E.D., DICKSON M.C., RODRIGUEZ A.C., GRIMWOOD J., SCHMUTZ J., MYERS R.M., BUTTERFIELD Y.S., KRZYWINSKI M.I., SKALSKA U., SMALLUS D.E., SCHNERCH A., SCHEIN J.E., JONES S.J. and MARRA M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

REFERENCE

2 (bases 1 to 2972)

Director MGC Project.

Direct Submission

Submitted (01-OCT-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: Howard Jacobs

cDNA Library Preparation: Express Genomics

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

cDNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) mcd@paxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

FEATURES

source

1..2972

/organism="Rattus norvegicus"

/mol_type="mRNA"

/db_xref="taxon:10116"

/clone="MGC:95109 IMAGE:7127425"

/tissue_type="Kidney, rat (Brown Norway)"

/clone_lib="NIH MGC_236"

/lab_host="DH10B"

/note="Vector: pExpress1"

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAP Plate: 184 Row: a Column: 6

This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

Location/Qualifiers

CDS

```
732..2444
/codon_start=1
/product="unknown (protein for MGC:95109)"
/protein_id="AAH83863.1"
/db_xref="GI:54035335"
/translation="MSRLRAKPKSLCKTEPLTSRVRSTLSVLKGIASQPSRLK
QLHNLGGCVCTTSSALLRNLSVTHPIKVLTSVONHVCCFSDCLFWALLCCNL
IENIRIGLTPPTAIKLNKYLGLSIVLKSEACSCRIPVDRSTHPLNLVHLSILTS
KPCAMLRKEIDHIGALILIKALLTIPESABRMVLGKSIIVPLKQVQVSTVLPLG
LLEAEVQLRLLPTQKSTLRVALFCASLSDGNAGEGLTVHVHYQVSLNVALEOL
LNLGRQLSDHDLVLCQKVIHPSLKQFLSEHQI IAI DRVGVTLMEPLSKVTGATPIG
SLYPIVSTYTGKVRGSRFGSKYFFHLLPNEATICSLLLCRNDTAEELKLTQCT
AMHVLQLTIKEPWVLKGGCVETHLAAVIRHKVHNEAEALVDDCSQAEHLHAEAF
CSALPSAAVSLHSDGEILIDMKYGHFWSCPADSASVGNWPTLSRCGGLYNSBEEL
SWSVLRSTYHPFAPQTCUPQAASGSVSNLTVDSFTAKSLGQVAVETANLILDLUSYVI
EDKN"
```

ORIGIN

```
Query Match      63.6%; Score 1571; DB 10; Length 2972;
Best Local Similarity 80.7%; Pred. No. 6.4e-265;
Matches 1898; Conservative 0; Mismatches 440; Indels 13; Gaps 5;

QY 1  TATGAGCCTTCGGAACTCTGTGAGAGACTACAAAAGTTTGGTGTGTTATGGTCCCTTTAGT 60
DB 117 TATGAGCTTTCAGAACTCTGTGAGAGAGTACAAAGTTCTGATGTGTTATGGTACCTTTAAT 176

QY 61  TGGGCTCATACATTTGGGGTGTGTACAGAAATCAAAAGACGCCCTGTTTCCAAATACCTAA 120
DB 177 CGGGTTTCATACATTTGGGGTGGCACAGAAATCAAAAGACGCCCTGTCTTCCAAATTCCT-- 234

QY 121 AAACGACGACATTCCTGACAGAGATGCTGGGACTTTCAAAATCTTCAGAGAGCCAAAT 180
DB 235 -AAGGACGATACCCCTGAGCTGGATAGTCTGGGGCTTGCAAGTCCACCAAGAGGCAACC 293

QY 181 CCAGGGAAAGTAGAGCGCTTGCAATCTTCAGGTAAAGAACAGCAGCTTTGAAATCTGAGCTTC 240
DB 294 CTAGGAGAGGAGCAAGCTTCGAGTTTTCAG--AAAGAGCAACTTTGAACTCGAACTTG 351

QY 241 ATATCGAAAGAGAGATGAAATACCACTTGGATTAGAAAGAACTGGCTTCTGTAGCT 300
DB 352 ATAATAAAAGAGAAATGAAATACCCAGTTGGATTGGAATAGAAAGAACTGGCTTCTGTGGCT 411

QY 301 GGGATATCTTCATAGTGTCCATCTTGGAACATATCTTTTCGAGAGGTCTGCAAGCAG 360
DB 412 GGGGTGACTTTTCGTAGGCGTCCATCTTGGAACGTACTTAATACAGAGAGCTCGGAGAG 471

QY 361 TCTGTAATAATTCAGTCTCAAAAGCAACAAAGAGATTATGAAGAGTGAAGTAAATAAAT 420
DB 472 TCTGTGAGGTCTCAGTCCAGAGACAAACAAAGGAATATTGAAGAGTGAAGCAAAATAAT 531

QY 421 ATTGTGAATTAATAATTTGTCATTAATCAATCTATGCTGATAGCTTCATAAAACATTGA 480
DB 532 ACTTGGAAATTAATA--TGCTCTTAATAATTCATCTATGCTGGTATTATGCTGAAA-----A 584

QY 481 ACTTTTGGATTTTATAGCCACATGCTGCATATTCATACTTTAAATTCCTAAAGATAATT 540
DB 585 CCACCTTGATTTTGTAGCTGGTAATGCTGATGTACAGACCTTTCATACCTGGGAGATAGTC 644

QY 541 TTTAATGTTTAAACAGTGAATAATGCAATAAATAGAAAATGTGGTTTACAAAATAAAAACG 600
DB 645 TTTAATGTTTCAATACATGACATTAATAATAAATGAAAACATGTGCAACAAATAAAGT 704

QY 601 GTCTTCACATGATTACACCTGAAGTAAGATGTCCTGTTTGGAGCTTAAGAAGCCATCAT 660
DB 705 GTCTTCACAACTTA-CACCTGAAGCAGAGTGTCTCGTTTGAAGCTTAAGAAGCCATCAT 763

QY 661 GTGTAGAGTGAACCACTGCAACACTGAGAGAGTTCAGGACCACTTTCTGTCCTGAAAG 720
DB 764 GTGTAAACTGAACCACTGCAACAGTGAAGAGTTCAGGTCTACATTTCTGTCTGAAAGG 823

QY 721 AATTGTGAACATCATGCTATGGCCCTTCAGGTAGGCTGAAGCAGCTGCAAAATGGCTTTGG 780
DB 824 AATCATAGCCTCATGCTATGGCCCTTCAGGGAGGCTGNAAGCAGCTGCACAATGGATTAGG 883
```

```
QY 781 AGGTTACGTGTGTACAACTCAGTCTCAGTCTGCTCAGTCACTTCTGCTCAGTCACTTTGGTCAACA 840
DB 884 AGGTTGTGTGTGTACAACTTACAGTCTCAGCCTGCTTGGAAACCTTTGGTCAACCCA 943

QY 841 TCCCAATTTTAAAGATTCCTGACAGCCTCATATACAGAAATCATGTGTCAAGCTTCAGTGATG 900
DB 944 TCCCAATTTTAAAGGTCCTAAACATCATCTGTGCGAGAAATCACGTGTCTGCTTCAGTGATG 1003

QY 901 TGGCTTATTACAGCTATTCTTTGCTGCAACCTGATTTGAAATGTTGAGAGATTAGGCTT 960
DB 1004 CGGCTTATTACAGCCATTCTGTGCTGCAACTTGATTTGAAATAATTCAGAGAAATAGGTTT 1063

QY 961 GACACCCACCACTGTCATATAGATTTAAATAAATCATCTTTTGTAGCTCTTTGATCATCAGTTA 1020
DB 1064 GACACCCCAACTGCTATTAAATTTAAATAAATACCTCTTGGGCTTCAGTATTCAGTTATCT 1123

QY 1021 CAAAGTCTGATACCTGCTGGTTGTGCAATCCAGTGGACCTTTAGTAGTACTCAGATCCTCT 1080
DB 1124 CAAAGTCTGAGGCTGTAGTTGTGCAATCCAGTGTGACTTCAGAAAGTACACATACCTCTCT 1183

QY 1081 TTGTTTGGTGGCTAGTATATTAAACAAAGTAAACCTGCTGTATGTCTCACAGAAAGAAAC 1140
DB 1184 CAACTTGGTACACAGTATCTTAAACAGCAACACAGCCTGCATGCTTACCAGAAAGAAAT 1243

QY 1141 AGAGATGTCAGTGTCTTGTATCTGAGAGCCCTTTTGTGCTTACAAATTCAGAAATGCTGA 1200
DB 1244 AGATCACATAGGTGTCTTTGATTTTGAAGAGCTTTTGTGCTTACAAATTCAGAAAGTGCAG 1303

QY 1201 AGGCCACATCAATTTAGGAAAGAGTTTAAATGTACTCTTTAAAGGTCAAGAGTTATAGA 1260
DB 1304 AGAGGCAATGGTTTTAGGGAAGAGTAAATTTGTTCCCTTTAAAGGCCAAACAGATTAACAG 1363

QY 1261 TTCCACTGTATTACCTGGGATACATTTGAAATGTGAGAAGTTCAGAAATTAAGTAAAGCTATT 1320
DB 1364 TTCTACTGTATTACCTGGACTTCTTATTGGAAGCATCAGAAAGTTCATTTAAGGAGATTAT 1423

QY 1321 ACCTATCAAAAATCAACTGCGCTCAAGTGGCACTCTTTGTACAACTTTATCCGGAGA 1380
DB 1424 ACCCAACAGAGTCAAGTACCTTTAGGTTGGCACTCTTCTGTGATCTTTATCTGAGAGA 1483

QY 1381 CACTTCTGCACTGGAGAAAGAACTGTGTGTGCTCAGTTATGGGGTTTCTCTTGAATAATGC 1440
DB 1484 TTTTCTTAATGCTGGAAGAGGAACTTGTGTGTGCTCAATTACCAGTGTCCCTTGAGAAATGC 1543

QY 1441 AGTCTTGAACAGCTGCTTAACTTAGGAAGGAGCTAATCAGTGACACGCTAGATGCTTGT 1500
DB 1544 AGTTTGTAGAGCAGTTGCTTAACTTAGGAAGACAGCTAGTCAAGTACCATGTAGATCTTGT 1603

QY 1501 CCTGTGCCAAAAGTTATACATCCATCTTTGAGAGCAGTTCTCTCATATATGCAATCTTAT 1560
DB 1604 CCTGTGCCAAAAGTATACACCCCATCTTTGAAACAGTTTCCCTCAGTGAAACATCAAAATTA 1663

QY 1561 TGCCATACAGAGAAATGGAGTGACTCTGATGGAACCCCTGACTAAATAGACAGGAACACA 1620
DB 1664 TGCCATACAGAGATTTGGAGTGACTCTGATGGAACCTCTGAGCAAGTGAACAGGAGCAAC 1723

QY 1621 GCCTATTGGATCCCTTAGGCTCAATATGCTCTTAATAGTTATGGAAGTGTGAAAGATGTGTG 1680
DB 1724 GCCTATTGGTTCCTTATACCCCAATAGTTTCTACTACTTATGGAAGTGTGAAAGATGTGCG 1783

QY 1681 CACTGCCAAAATTTGGCTCCCAACATTTTTTTCATCTTATCTCTATTAAGCAACAATCTG 1740
DB 1784 CTCTGCAAGATTTGGCTCCAAATATTTTTTTCATCTTCTCTTAATAGGGGACCATCTG 1843

QY 1741 CAGCTTGTCTCTGCAACAGAAATGACACTGCGCTGGGATGAGCTGGAAGCTCACCTGTCA 1800
DB 1844 CAGCTTGTCTTCTGCGAGCAAGAAATGACACTGCGCTGGGAGAGCTGGAAGCTCACCTGTCA 1903

QY 1801 GACGGCACTGCATGCTCTGCAAGTTAACTCAAGGAACCATGGGCTTTTGTGGGAGGTGG 1860
DB 1904 AACGCAATGCATGCTCTTGCAAGTTAACTCAAGGAACCATGGGCTTTTATTTGGGAGGTGG 1963
```


Db 480 TCTGTGAGTCTGAGGCCAGAGCAGACAAAGAAATATTTGAAGAATGAAGCAAAATAAAG 539
Qy 421 ATTTGGAATTAATAATTTGTCATTAATAATCATTTCTATGCTGATAGTTCATTAACAATTGA 480
Db 540 ACTTTGGATTAATAA--TGTCATTAATAATTAATCTATGCTGGTTATTTGCTGAAACCCACTT 597
Qy 481 ACTTTTGTGATTTTATAGCCACCAATGCTGCATATTCATATCTTTAAATTTCTTAAGAATAAT 540
Db 598 GAATTTGTAGTGGT-----AATGCTGCATGTCAGAC--TCATACCTGGAGACCAAGTC 649
Qy 541 TTTAATGTTAAACGTCGTAATAATGCAATAAATAAGAAAATGTGGTTTACAAAATAAAAAAG 600
Db 650 TTTAATGTTCAACACATGACAGTGCATATAAATAAATGAAGCCATATGAACAAACTAAAGCT 709
Qy 601 GTCTTCACATGTTACACCTGAAAGTAAGATGCTCGTTTGGAGAGCTAAGAAGCCATCAT 660
Db 710 GCCTTCACAAAGTTA-CACCTGAAGCAGGATGTCTCGTTTAGAAGCTTAAGAAGCCATCGTT 768
Qy 661 GTGTAGAGTGAACCACTGACAACTGAGAGAGTGCAGGACCACACTTTCTGCTCTGAAAAG 720
Db 769 GTGTAAACTGAACCAATGACAAAGTGAGAAAGTCAAGTCCACACTTTCTGCTCTGAAAG 828
Qy 721 AATTTGAACATCATGCTATGGCCCTCAGGTAGGCTGAAGCAGCTGCACAAATGGCTTTGG 780
Db 829 AGTCATAGCTCATGCTATGGCCCTTCAGGGAGCTGAAGCAGCTACACAAATGGATTGG 888
Qy 781 AGTTACGTGTGTACAACTCAGCTCCTCAGTCTCAGTCTGCTCAGTCACTCTTTGGTGCACAC 840
Db 889 AGGTTGTGTGTACAACTCAGTCTCAGTCTCAGGCTCTGCAACCTTTTCAGTCACTCA 948
Qy 841 TCCATTTTAAAGATCCTGACAGCTCCATACAGAAATCATGTGTCAAGCTTCAGTGATG 900
Db 949 TCCCGTATTAAGATCCTAACGTCATCCGTGCAGAAATCAGTGTCTGCTTCAGTGACTG 1008
Qy 901 TGGCTTAATCAGAGCTAATCTTTTGTGCAACCTGATTTGAAAATGTTTCAGAGATTAGGCT 960
Db 1009 CGGCTTAATCAGAGCTAATCTGTGCTGCAACTTTGATTGAAAATATTTCAAAGACTAGATT 1068
Qy 961 GACACCACACGTGCTAATTAAGTAAATAAACAATCTTTTGGTCTTTGGCATCAGTTATCT 1020
Db 1069 GACACCCGAACCTGCTAATTAATAATAAATACCTCTTGGAGTCTCTGTACCAGTTATCT 1128
Qy 1021 CAAGTCTGATACCTGTGTTGTCGAATCCAGTGGACTTTTAGTGTGCTCAGATCCTCT 1080
Db 1129 CAAGTCTGAAGCTGTAGTTGTCGAATCCAGTTGACTTCAGAAATACATACCTTCT 1188
Qy 1081 TTGTTTGGTGGTAGTATTAACAAGTAAACCTGCTGTATGCTCACAGAAAGAAAC 1140
Db 1189 CAGCTTGGTACACAGTATCTTAACAAGCAAAACAGCTGTATGCTCACAGAAAGAAAC 1248
Qy 1141 AGAGCATGTGCTGCTTGTATCCTGAGAGCTTTTGTGCTTACAAATTCAGAAAATGCTGA 1200
Db 1249 AGATCATAGTGTGCTTGTATTTGAAAGCTTTTACTTACAAATTTCCAGAAAGTACAGA 1308
Qy 1201 AGSCCACAATCTTTTAGGAAAGTAAATTTGACCTTTTAAAGGTCAAGAGTTATAGA 1260
Db 1309 AGAAGAAATGTTTATAGGAAGAGTAAATTTGTTCCCTTAAGGGCCAAAGAGTTACAGA 1368
Qy 1261 TTCCACTGTATTAACCTGGGATCTCAATGAATGTGAGAAATGCTCAATTAATGAGGCTATT 1320
Db 1369 TTCTACTGTATTAACCTGGACTACTCATTTGAAGCATCAGAAGTTCAATTAAGGAGGTTATT 1428
Qy 1321 ACCTATCAAAAATCAACTGCCCTCAAGGTGGCACTTTTGTGACAACTTTATCCGGAGA 1380
Db 1429 ACCACTCAAAAGGCAAGTGGCCCTTAGGGTGGCACTCTCTGTACATCTTTATCTGGAGA 1488
Qy 1381 CACTTCTGACACTGGAGAGGAACCTGTGTGGTGCAGTTATGGGTTTCTCTTGAAATGC 1440
Db 1489 TTTTCTAATGCTGGAGAGGAGTGTGTGGGCCAATTACCAAGTGTCCCTTGAGATGC 1548
Qy 1441 AGCTTTGGACCACTGCTTAACTAGGAAGGCAAGCTTAATCAGTGACCACTAGATCTTGT 1500
Db 1549 AGTTTTAGAGCAGTTGCTTAATCTAGGAAGGCGACTAGTCACTGATCATGTAGACCTTGT 1608

Qy 1501 CCTGTGCCAAAAGTTATACATCCATCTTTGAAGCAGTTTCTCAATATGCATCGTATTAT 1560
Db 1609 CCTGTGCCAAAAGTAAATCCACCCATCTTTGAAAACAGTTCTTCAGTGAGCGCAGCTTAT 1668
Qy 1561 TGCCATAGACAGAAATGGAGTGACTCTGATGGAAACCCCTGACTAAATAATGACAGGAACACA 1620
Db 1669 GGCCATACAGAGATTGGGTGACTCTGATGGAATCTCTAAGCAAGTGAAGAGGACAAAC 1728
Qy 1621 GCCTATTGGATCCCTAGGCTCAATATGCTTAATAGTCTTAATAGTGTGAAGATGTGTG 1680
Db 1729 GCCTATTGGTCTCTTAAACCCCAATAGTTCTTACTACTTATGGAAGTGTGAAGATGTGTG 1788
Qy 1681 CACTGCAAAATTTGGCTCCCAACATTTTTTTCATCTTATCTTAATGAAGCAACATCTG 1740
Db 1789 CTCTCAAGATTTGGCTCCAAACATTTTTTTCATCTTCTTCAATGAGGCCACGTCTG 1848
Qy 1741 CAGCTTGTCTCTCTCAACAGAAATGACATGCTCCCTGGATGAGCTGAAGCTCACGTGTCA 1800
Db 1849 CACCTTGTCTCTCTCAGCAGAAATGACACCGCTGGAGGAGCTGAAGCTCACATGTCA 1908
Qy 1801 GACGCCACTGCATGCTCTGCACTCAAGGAAACATGGGCTTTGTTGGGAGGTGG 1860
Db 1909 AACAGCAATGCAGCTCTTGCAAGTAAACAATCAAGGAAACCGTGGGTTTTATTGGGAGGTGG 1968
Qy 1861 CTGTACTGAAACTCATTTGGCTGCAATATACAGACACAAGACTCAACAGGCCCAAGAAAG 1920
Db 1969 CTGTACAGAAACACACTTGGCTGCAATATGTACAGACAAAGGTTTCATCAGGAGGAGAGC 2028
Qy 1921 CATTTCTCAAGATGATGATGATCTCAAAACAGAACTTCAATTTAAATGCTGAAGCAATTTG 1980
Db 2029 TATTGTCAGAGATGATGGGTGACTCAGGCAAAAGTGCATGTTGCTGTGAAGCAATTTG 2088
Qy 1981 CAGTCCCTTAGAATCTGTTGGTCTTTTAGAAACATGATGAGGTGAAAATCTCACTGA 2040
Db 2089 CAGTGTCTGGAGTCCGTTGCTGGCTCTTTGGAAACATGATGGTGGTGAATCCTCATTTGA 2148
Qy 2041 CATGAAGTATGACACCTTTGGTCAAGTTCAGGACATCTCCCTGTTGCTTAAGTGGCC 2100
Db 2149 CACGAAGTATGACACCTTTGGTCTGCTCAAGCAGATCTGCTCTGTTGTTAACTGGTC 2208
Qy 2101 AGATTTGCTTTTCAAGTGTGGCTGTGGATTATACAAATAGCCAGGAAGAACTCAACTGGTC 2160
Db 2209 AGATACGCTGTCAACGTTGGCTGTGGTTTGTACAAAGCAGGAAGAGCTCAGCTGGTC 2268
Qy 2161 TTTCTTAAGAAAGCAACGTCGTCATTTGTGTCACAAAGCTGCCCTTCCACATGAAGCTGT 2220
Db 2269 TGTCTTAAGAAAGTACTTATCATCTCTTTTGACCAACAAACCTGCCCTTCCACAGGAGCTTT 2328
Qy 2221 GGGCTCAGCCAGCAACCTGACCTTGGACTGTTTGAAGTCAAGAGCTTAGTGGCTACAGGT 2280
Db 2329 GGGCTCAGCCAGTAACTGACTGTGGACTGCTTCACTGCAAGCTGAGTGGCTTACAGT 2388
Qy 2281 GGCTGTAGAGACAGCAATTTGATTTGGGATCTTTTATATGTTTGAAGATAAAAACTA 2340
Db 2389 GGCTGTAGAGACAGCAATTTGATTTAGATCTTTTATATGTCATTTGAAGATAAAAACTA 2448
Qy 2341 AGAAGATAGCATG 2353
Db 2449 ACATAAGAGAATG 2461

RESULT 11
AF254074
LOCUS
DEFINITION
Mus musculus MKKS protein (Mkks)
ACCESSION
AF254074
VERSION
AF254074.1
KEYWORDS
GI:8164187
Mus musculus (house mouse)
ORGANISM
Mus musculus

AF254074 2699 bp mRNA linear ROD 02-JUN-2000
Mus musculus MKKS protein (Mkks) mRNA, complete cds.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 2699)
 AUTHORS Stone, D.L., Slavotinek, A., Bouffard, G.G., Banerjee-Basu, S., Baxevanis, A.D., Barr, M. and Biesecker, L.G.
 TITLE Mutation of a gene encoding a putative chaperonin causes McKusick-Kaufman syndrome
 JOURNAL Nat. Genet. 25 (1), 79-82 (2000)
 MEDLINE 20264373
 PUBMED 10802661

REFERENCE 2 (bases 1 to 2699)
 AUTHORS Slavotinek, A.M., Biesecker, L.G. and Stone, D.L.
 TITLE Direct Submission
 JOURNAL Submitted (10-APR-2000) NHGRI, NIH, 49 Convent Drive, Bethesda, MD 20892, USA

FEATURES
 source 1..2699
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 gene 1..2699
 /genes="Mkks"
 /notes="putative chaperonin"
 CDS 10802661
 /product="MKKS protein"
 /protein_id="AAF73965.1"
 /db_xref="GI:8164188"
 /translation="MSRLBAKPSLCKTEPLTSEKVRSTLSVLKGVIASCYGSPRLK
 OLHNGIGCVYTTSSQALLNLNVTHPVVKILITSSVQNHVSCFSDGCLPTAILCNL
 IENIQRLDLTPATAIKLNYLLSLCTSYLSEKACSCRIIPDRFSTHTFSLVHSILTS
 KPCMTQRRETHIGALIIKAPLITIPESTBERVLGKSIIIVPLKQVRVDSVLPGL
 LIEASVQRLRLPTQKAGLRVALFCTSLSDGFSNAGGVVVAHVQVLENAVLQOL
 LNLRLRLVDVLCQKVIHPSLRQFSPSRHVMADIRGVTLMESLSKVKCATPIG
 SLNPIVSTYGVSKVCSARFGSKHFPFLHPNEATVCTLLCSRNDTAEELKLTCT
 AMHVLQLTIKEPVLVGGCTETHLAAVVRHKVHAEAIVRDDGCTCOKLHVAAEF
 CSALESVAGSLHEDGELIDTKYGHLSQDASVGVNWSDTLSRGCGLNRSBEL
 SWSVLRSTYHPFAPQCLPQALGASASNLTVDCFTAKUSGLQVAVETANLILDLSTVI
 EDKN"

ORIGIN
 Query Match 62.5%; Score 1545.8; DB 10; Length 2699;
 Best Local Similarity 80.3%; Pred. No. 1.6e-260;
 Matches 1890; Conservative 0; Mismatches 447; Indels 16; Gaps 6;

QY 1 TATGAGCCTTCGGAACTTTGTGAGAGACTACAAAGTTTGTGTTGTTATGTCCTTACT 60
 DB 157 TATGAGCTTCAGAACTCTGGAGAGAGTACAAAGTTCTGATGTTATGTTACCTAAT 216
 QY 61 TGGGCTCATACATTTGGGGTGTACAGAAATCAAAGCAGCCCTGTTTCCAAATACCTAA 120
 DB 217 TGGGTTTCATACATTTGGGGTGGCACAGAAATCAAAGCAGCCCTGTTCCAAAGTTCCT-- 274
 QY 121 AAACGACGACATTCCTGAGCAGATAGTCTGGGACTTTTCAATCTTCAGAGAGCCAAAT 180
 DB 275 -AAGGACGATACACATGAGCTGGATAGTCTGGGACTTGCAGGTCACCAAGAGGCCAAAC 333
 QY 181 CCAGGGGAAGTAGCAGCTTGCATCTTCAAGGTAAAGAGCAGCTTTGAAATCTGAGCTTC 240
 DB 334 CTAGGAGAGTAGCCATCTTGAGTTTTCAG--AAAGAGCACTTTGAAATCTGAACTTG 391
 QY 241 ATATCGAAAGAGAGATGAAATATACAGTTGGATTAGAAAGACTGGCTTCTTGTAGCT 300
 DB 392 ATAATAAAGAGAAATGAAAGCCCAAGTGGATTAGAAAGAACTGGCTTCTTGTGGCT 451
 QY 301 GGGATATCTTCATAGGTGTCATCTTGGAACTATCTTTTGCAGAGGTCTGCAAGAGCAG 360
 DB 452 GGGGTGACTTTTCATAGGCGTCCATCTTGGAACTATCTTTATACAGAGGTTCGCAAGAG 511
 QY 361 TCTGTAAATTTTCAGTCTCAAAGCAAAACAAAGAGATTTGAAGAGTGAAGTAAATAAT 420
 DB 512 TCTGTAGGTCTGAGGCCAGAGCAGACAAAGAGATTTGAAGATGAAGCAAAATAAG 571
 QY 421 ATTTGGAATTTACTAATTTGTCAATTAATCAATTTCTATGCTGATGATGCAATTAAG 480

DB 572 ACTTTGGATTATTA--TGTCAATAAATAATCTATGCTGGTTATTTGTCTGAACCACTT 629
 QY 481 ACTTTTGTGATTTTATAGCCACAAATGTCATATTCATATCTTTAAATTCCTTAAAGAAATTT 540
 DB 630 GATTTTGTAGTGGT-----AATGCTGATGTGACAGC--TCATACCTTGGAGCCAGTC 681
 QY 541 TTTAATGTTAAACGCTGATAATGCAATAAATAAGAAAAATGTGGTTTACAAAATAAAAAACG 600
 DB 682 TTTAATGTTTCAAAACACATGACAGTGCATAATAAATAAGAACCATATGACAAACTAAAGCT 741
 QY 601 GTCTTCACCTAGTATACACCTGAAGTATGATGCTCTGTTTGGAGCTTGAAGCCATCATTT 660
 DB 742 GCCTTTCACAAAGTTA--CACCTGAAGCAGGATGTCTGGTTAGAGCTTGAAGCCATCGTT 800
 QY 661 GTGTAAGAGTGAACCACTGACAACTGAGAGAGTTCAGGACCAACACTTTTCTGTCTTGAAG 720
 DB 801 GGTGTAACCTGAACCACTGACAACTGAGAGAGTTCAGGACCAACACTTTTCTGTCTTGAAG 860
 QY 721 AATTGTAAACATCATGCTATGGCCCTTCAGGTAGGCTGAAGCAGCTGCACAAATGGCTTTGG 780
 DB 861 AGTCATAGCCTCATGCTATGGCCCTTCAGGGAGGCTGAAGCAGCTACACAATGGATTGG 920
 QY 781 AGTTACGTGTGTACAACTCAGCTCTCAGCTCTGCTCAGTCTGCTTGGTGTACACA 840
 DB 921 AGGTTGTGTGTACACAACCTCAGCTCTCAGCCCTGCTTCGAAACCTTTTCAGTCAACCCA 980
 QY 841 TCCCAATTTTAAAGATCCTGACAGCCCTCCATACAGAAATCATGTGTCAGCTTCAAGCTTCAGTCAT 900
 DB 981 TCCCGTATTAAGATCCTTAACCTCATCCGTGAGAAATCAAGTCTGCTCTGCTTCAAGTACTG 1040
 QY 901 TGGCTTATTTACAGCTATTTCTTTGTGCAACCTGATGAAATGTTTCAAGATTAAGCTT 960
 DB 1041 CGGCTTATTTACAGCACTTTCTGCTGCACTTGTGAAATATTTCAAAGACTAGATT 1100
 QY 961 GACACCCACCATGTCATATGATTAATAATAAATCATCTTTTGTGAGTCTTTGCAATCAGTTATCT 1020
 DB 1101 GACACCCGCACTGCTATTAATAATAATAACCTCTTGTGAGTCTCTGTACCAAGTTATCT 1160
 QY 1021 CAAGTCTGATACCTGTTGTCGAATCCAGTGGACTTTTGTAGTACTCAGATCCTCTCT 1080
 DB 1161 CAAGTCTGAAGCTTGTGTTGTCGAATCCAGTGGACTTCAAGAGTACACATACCTTCTCT 1220
 QY 1081 TTGTTTGGTGGCTAGTATATTAACAAGTAAACCTGCTGTATGCTCACCAGAAAGGAAAC 1140
 DB 1221 CAGCTTGGTACACAGTATCTTAACAGCAACAGCCTGTATGCTCACCAGAAAGGAAAC 1280
 QY 1141 AGAGCATGTCAGTCTTTGATCCTCAGAGCCCTTTTGTCTTACAATTCAGAAATGCTGA 1200
 DB 1281 AGATCACATAGGTGCTTTGATTTTGAAGCTTTTGTACTTACAATTCAGAAAGTACAGA 1340
 QY 1201 AGGCCACATCATTTTAGGAAAGAGTTTAATGTTACTCTTTAAAGGTCAAGAGTTATAGA 1260
 DB 1341 AGAAGCAATGGTTTATAGGAAAGAGTAAATGTTCCCTTAAAGGGCCAAAGAGTTACAGA 1400
 QY 1261 TTCCACTGTATTACCTGGGATCTCATTTGAAATGTTCAGAAATTCAGAAATTAATGAGGCTATT 1320
 DB 1401 TTCTACTGTATTACCTGGACTACTCATTTGAGCATCAGAGTTCAATTAAGAGGTTATT 1460
 QY 1321 ACCTATCAAAAATCAACTGCCCTCAAGTGGCCTCTTTTGTACAACCTTTTATCCGGAGA 1380
 DB 1461 ACCCACTCAAAAGGCAAGTGGCCTTAGGTTGGCACTCTTCTGTACATCTTTTCTGGAGA 1520
 QY 1381 CACTTCTGACACTGGAGAGGAACTGTTGGTGTGAGTTATGGGTTTCTCTTGAANAATGC 1440
 DB 1521 TTTTCTTAAATGCTGGAGAGGAGTGGTGGTGGCCCATTAACCAAGTGTCCCTTTGAGAATGC 1580
 QY 1441 AGCTCTTGACCACTGCTTAACTAGGAGGAGGAGCTTAATCAGTGCACCACTGATCTTCTGT 1500
 DB 1581 AGTTTGTAGCAGTTGCTTTAATCTAGGAGGAGGCACTAGTCACTGATCATGTAGACCTTGT 1640
 QY 1501 CTTGTGCCAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGCAATCGTATTAT 1560

Db 1641 CCTGTGCCAAAAGTAATCCACCCTCTTTGAGACAGTCTTCTTCAGTGAGCGCCACGTTAT 1700
QY 1561 TGCATATAGACAGAAATGGAGTGAGTCTGTATGGAACCCCTGACTAAATGACAGGACACACA 1620
Db 1701 GGCCATCGACAGAGTTGGGGTGAATCTGTATGGAATCTCTAAGCAAAAGTCACAGGACCAAC 1760
QY 1621 GCGTATTGGATCCCTAGGCTCAATATGCTCTTAATAGTATGAGTGAAGTGAAGAGATGCTGG 1680
Db 1761 GCGTATTGGTCTCTTAACCCCAATAGTTTCTACTTATGAGAGTGAAGAGATGCTGG 1820
QY 1681 CACTGCAAAATTTGGCTCCCAACATTTTTTTCATCTTATCTTCTTAATGAAGCAACAATCTG 1740
Db 1821 CTCTGCAAGATTTGGCTCCCAACATTTTTTTCATCTTCTTCTTCTTCTTCTTCTTCTG 1880
QY 1741 CAGTTGCTTCTCTGCAACAGAAATGACACTGCTGCGGATGAGCTGAAGCTCACTGTGCA 1800
Db 1881 CACCTTGTCTTCTGCGAGCAAAATGACACCGCTGGGAGAGCTCAAGCTCACATGTCA 1940
QY 1801 GAGGCACTGCATGCTCCTGAGTTAACTCACTCAAGGAACCATGGCTTTGTTGGGAGGTGG 1860
Db 1941 AACAGCAATGACAGTCTTTCAGTTTAACTCAAGGAACCGTGGGTTTATTGGGAGGTGG 2000
QY 1861 CTGTACTGAACATCACTTTGGCTGCATATATCAGACACAGACTCAACAGCCCAAGAAAG 1920
Db 2001 CTGTACAGAAACACACTTGGCTGCATATGTCAGACACAGGTTTCATCAGGAGCAGAGC 2060
QY 1921 CATTCTCAAAGATGATGAATGATCAACAGAACTTTCAATTAATGCTGAAGCATTTTG 1980
Db 2061 TATTGTACAGATGATGGGTGACTCAGGCAAAAGCTGCAATGTTGCTGCTGAAGCATTTTG 2120
QY 1981 CAGTGCCTTAGAATCTGTTGGCTCTTTAGAACATGATGAGTGAATTTCTCACTCA 2040
Db 2121 CAGTGTCTGAGAGTCGGTTCGCTCTTTGGAACATGATGGTGAATCCTCATTTGA 2180
QY 2041 CATGAAGTATGACACACTTTGGTTCAGTTTCAGGAGATTTCTCCTGTGTGCTAACTGGCC 2100
Db 2181 CACGAGTATGACACACTTTGGTTCGCTGTCAGCAGATTTCTGCTGTGTGTTAACTGGTC 2240
QY 2101 AGATTGCTTTCACAGTGGCTGTGGATATACATATGACAGGAGAACTCAACTGGTTC 2160
Db 2241 AGATAGCTGTACAGGTGGCTGTGGTGTGTGTACACAGCCAGGAGAGCTCAGCTGGTC 2300
QY 2161 TTTCTTAAGACACACAGTCGCTCCATTTGTCGCCACAAAGCTGCCTTCCACATGAAGCTGT 2220
Db 2301 TGTCTTAAGAGTACTTATCATCTTTTGACACAAACCTGCCTTCCACAGGAGCTTT 2360
QY 2221 GGGCTCAGCAGCAACCTGACCTTGGACTGTTTGAAGTCAAGCTTAGTGGCTACAGCT 2280
Db 2361 GGGCTCAGCCAGTAACCTGACTGTGGAGTCTTCACTGCCAAGCTGAGTGGCTTACAGGT 2420
QY 2281 GCGTGTAGACAGCCAAATTTGATTTGGGATCTTTTCATATGTTATTTGAAGATAAAACTA 2340
Db 2421 GCGTGTAGACAGCCAAATTTGATTTAGATCTTTTCATATGTTATTTGAAGATAAAACTA 2480
QY 2341 AGAGATAGCAATG 2353
Db 2481 ACATAAGAGATG 2493

RESULT 12
LOCUS HS931K24 74813 bp DNA linear PRI 15-MAR-2001
DEFINITION Human DNA sequence from clone RPS-931K24 on chromosome 20p12.
Contains ESTs, STSs, GSSs and two CpG islands. Contains a novel pseudogene, the MKKS gene for McKusick-Kaufman syndrome protein and the 5' end of a novel gene, complete sequence.
ACCESSION AL034430
VERSION AL034430.4 GI:11322731
KEYWORDS HIG; CpG island; MKKS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

1 (bases 1 to 74813)
Bates,K.
Direct Submission
Submitted (12-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Nov 23, 2000 this sequence version replaced gi:9795189.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
Chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
IMPORTANT: This sequence is not the entire insert of clone
RPS-931K24 It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RPS-931K24 is at 74813 in this
sequence. The true left end of clone RPl1-254M13 is at 26700 in
this sequence. The true right end of clone RPS-1068F16 is at 100 in
this sequence. This sequence was finished as follows unless
otherwise noted: all regions were either double-stranded or
sequenced with an alternate chemistry or covered by high quality
data (i.e., phred quality >= 30); an attempt was made to resolve
all sequencing problems, such as compressions and repeats; all
regions were covered by at least one plasmid subclone or more than
one M13 subclone; and the assembly was confirmed by restriction
digest. RPS-931K24 is from the library RPl1-5 constructed by the
group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2.

FEATURES

source

Location/Qualifiers

1..74813
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="p12"
/clone="RPS-931K24"
/clone_lib="RPCI-5"
2..154
/note="AluJb repeat: matches 1..153 of consensus"
367..387
/note="BC200 repeat: matches 126..145 of consensus"
388..500
/note="FLAM C repeat: matches 12..124 of consensus"
1066..1263
/note="MER46C repeat: matches 103..304 of consensus"
1626..1922
/note="AluSp repeat: matches 1..299 of consensus"
complement(1821..2342)
/note="match: GSS: Em:AQ281134"
2118..2257
/note="MER5A repeat: matches 23..189 of consensus"
2282..2403
/note="MER5A repeat: matches 9..154 of consensus"
complement(2331..2443)
/note="match: GSS: Em:AF156717"
2788..2947
/note="FRAM repeat: matches -2..157 of consensus"
3484..3577
/note="MER53 repeat: matches 89..189 of consensus"
3600..4195
/note="LTR29 repeat: matches 3..619 of consensus"
complement(4233..4699)

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

misc_feature

repeat_region

repeat_region

misc_feature

repeat_region

repeat_region

repeat_region

misc_feature

Db	47002	GCAATAAATAGAAAAATGTGGTTTACAAATAAAAAACGGTCTTCACTAGTTACCACCTGA	46943	LOCUS	BD276497	1274 bp	DNA	linear	PAT 17-JUL-2003
QY	623	AGTAAGATGTCGTGTTTGAAGCTAAGAAGCCATCAATTGTGTAAGAGTGAAACCACTGACA	682	DEFINITION	143 Human Secreted Proteins.				
Db	46942	AGTAAGATGTCGTGTTGGAAGCTAAGAAGCCATCAATTGTGTAAGAGTGAAACCACTGACA	46883	ACCESSION	BD276497				
QY	683	ACTGAGAGAGTCAGGACCACTCTCTGTCCTTGAAAGAATTTGTAACATCATGCTATGGC	742	VERSION	BD276497.1	GI:33086265			
Db	46882	ACTGAGAGAGTCAGGACCACTCTCTGTCCTTGAAAGAATTTGTAACATCATGCTATGGC	46823	KEYWORDS	JP 2002543836-A/169.				
QY	743	CCCTCAGGTAGGCTGAAGCAGCTGCACAATGGCTTTGGAGGTTACGTGTGTACAACCTCA	802	SOURCE	Homo sapiens (human)				
Db	46822	CCCTCAGGTAGGCTGAAGCAGCTGCACAATGGCTTTGGAGGTTACGTGTGTACAACCTCA	46763	ORGANISM	Homo sapiens				
QY	803	CAGTCCTCAGCTGCTGCTCAGTCACCTTTTGGTCACACATCCCATTTTAAAGATTCCTGACA	862	REFERENCE	Soppet,D.R., Florence,K.A., Duan,R.D., Bias,C.E., Komatsoulis,G.A.,				
Db	46762	CAGTCCTCAGCTGCTGCTCAGTCACCTTTTGGTCACACATCCCATTTTAAAGATTCCTGACA	46703	AUTHORS	Young,P.E., Moore,P.A., Steven, Ruben,M. and Rosen,C.A.				
QY	863	GCCTCCATACAGNAATCATGTGTCMAAGCTTCAGTGAATGTGGCTTATTTCACAGCTATTCTT	922	TITLE	143 Human Secreted Proteins				
Db	46702	GCCTCCATACAGNAATCATGTGTCMAAGCTTCAGTGAATGTGGCTTATTTCACAGCTATTCTT	46643	JOURNAL	Patent: JP 2002543836-A 169 24-DEC-2002;				
QY	923	TGCTGCAACTGATTCAAAATGTTACAGATTAGGCTTGACACCCACCACTGTCAATTAGA	982	COMMENT	OS Homo sapiens				
Db	46642	TGCTGCAACTGATTCAAAATGTTACAGATTAGGCTTGACACCCACCACTGTCAATTAGA	46583		PN JP 2002543836-A/169				
QY	983	TTAAATAAACATCTTTTGAGTCCTTTGTCATCAGTTATCTCAAGTCTGATACCTGTGTTGT	1042		PD 24-DEC-2002				
Db	46582	TTAAATAAACATCTTTTGAGTCCTTTGTCATCAGTTATCTCAAGTCTGAGACCTGTGTTGT	46523		PF 11-MAY-2000 JP 2000618448				
QY	1043	CGAATCCCACTGAGTACTGATGACTCAGATCCCTCTTTGTTGGTGGGTAGTATATTA	1102		PR 13-MAY-1999 US 60/134068				
Db	46522	CGAATCCCACTGAGTACTGATGACTCAGATCCCTCTTTGTTGGTGGGTAGTATATTA	46463		PI daniel r soppet,kimberly a florence, roxanne d duan,charles e				
QY	1103	ACAAGTAAACCTGCCCTGTATGCTCACAGAAAGGAAACAGAGCATGTCAGTCTTTGATC	1162		PI bias,				
Db	46462	ACNAGTAAACCTGCCCTGTATGCTCACAGAAAGGAAACAGAGCATGTCAGTCTTTGATC	46403		ruben,craig a rosen				
QY	1163	CTGAGAGCCTTTTGGCTTACAATTCAGAAATGCTGGAAGGCCACATCATTTTAGGAAAG	1222		CC				
Db	46402	CTGAGAGCCTTTTGGCTTACAATTCAGAAATGCTGGAAGGCCACATCATTTTAGGAAAG	46343		FH Key				
QY	1223	AGTTTAATGTACCTTTAAAGGTCAAAGAGTTATAGATTCACCTGTTATTCCTGGATA	1282	FEATURES	Location/Qualifiers				
Db	46342	AGTTTAATGTACCTTTAAAGGTCAAAGAGTTATAGATTCACCTGTTATTCCTGGATA	46283	source	1. .1274				
QY	1283	CTCATTTGAAATGTCAGAAAGTTCAATTAATGAGGCTATTACCTATCAAAAAATCAACTGCC	1342	ORIGIN	/organism="Homo sapiens"				
Db	46282	CTCATTTGAAATGTCAGAAAGTTCAATTAATGAGGCTATTACCTATCAAAAAATCAACTGCC	46223	Query Match	50.7%; Score 1253.8; DB 6; Length 1274;				
QY	1343	CTCAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGACACTTCTGACACTCGAGAAGGA	1402	Best Local Similarity	99.8%; Pred. No. 2.1e-209;				
Db	46222	CTCAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGACACTTCTGACACTCGAGAAGGA	46163	Matches 1266; Conservative	0; Mismatches 2; Indels 1; Gaps 1;				
QY	1403	ACTGTGGTGGTCAGTTATGGGGTTCTCTTGAATGTCAGTCTTGACCAGCTGCTTAAC	1462	QY	1156	TTTGATCCTGAGAGCCTTTTGGCTTACAAATTCACAAAATGCTGAAGGCCACATCATTTT	1215		
Db	46162	ACTGTGGTGGTCAGTTATGGGGTTCTCTTGAATGTCAGTCTTGACCAGCTGCTTAAC	46103	Db	1	TTTGATCCTGAGAGCCTTTTGGCTTACAAATTCACAAAATGCTGAAGGCCACATCATTTT	60		
QY	1463	CTAGGAAGGCAGCTAATCAGTGACCAGTAGATCTTGTCTGTGCCAAAAAGTTATACAT	1522	QY	1216	AGGAAGAGTTTTAATTTGTACCTTTTAAAGGTCAAAAGGTTATAGATTCCACCTGTATTACC	1275		
Db	46102	CTAGGAAGGCAGCTAATCAGTGACCAGTAGATCTTGTCTGTGCCAAAAAGTTATACAT	46043	Db	61	AGGAAGAGTTTTAATTTGTACCTTTTAAAGGTCAAAAGGTTATAGATTCCACCTGTATTACC	120		
QY	1523	CCATCTTTGAAGCAGTTTCTCAATATGCATCGTATTATTGGCCATAGACAGAAATGGAGTG	1582	QY	1276	TGGGATACTCATTTGAAATGTCAGAAAGTTCAAATTAATGAGGCTATTACCTATCAAAAAATC	1335		
Db	46042	CCATCTTTGAAGCAGTTTCTCAATATGCATCGTATTATTGGCCATAGACAGAAATGGAGTG	45983	Db	121	TGGGATACTCATTTGAAATGTCAGAAAGTTCAAATTAATGAGGCTATTACCTATCAAAAAATC	180		
QY	1583	ACTCTGATGGACCCCTGACTAAAATGACAGGAACACAGC	1622	QY	1336	AACTGCCCTCAAGGTGGGCACTCTTTTGTACAACTTTATCCGGAGACACTTCTGACACTGG	1395		
Db	45982	ACTCTGATGGACCCCTGACTAAAATGACAGGAACACAGC	45943	Db	181	AACTGCCCTCAAGGTGGGCACTCTTTTGTACAACTTTATCCGGAGACACTTCTGACACTGG	240		
				QY	1396	AGAAGGAACTGTGGTGGTCAGTTATGGGGTTTCTCTGAAAATGCACTCTTGACACAGCT	1455		
				Db	241	AGAAGGAACTGTGGTGGTCAGTTATGGGGTTTCTCTTGAATGTCAGTCTTGGAACAGCT	300		
				QY	1456	GCTTAACTCTAGGAAGGCAGCTAATCAGTGACCACTAGATCTTGTCTGTGCCAAAAAGT	1515		
				Db	301	GCTTAACTCTAGGAAGGCAGCTAATCAGTGACCACTAGATCTTGTCTGTGCCAAAAAGT	360		
				QY	1516	TATACATCCATCTTTGAAGCAGTTTCTCAATATGCATCGTATTATTGGCCATAGACAGAT	1575		
				Db	361	TATACATCCATCTTTGAAGCAGTTTCTCAATATGCATCGTATTATTGGCCATAGACAGAT	420		
				QY	1576	TGGAGTGACTCTGATGGAAACCCCTCAGCTAAATATGACGAAACACAGCCTTATTGGATCCCT	1635		
				Db	421	TGGAGTGACTCTGATGGAAACCCCTCAGCTAAATATGACGAAACACAGCCTTATTGGATCCCT	480		
				QY	1636	AGGCTCAATATGTCTCTTAATAGTTATGGAAAGTGTGAAAGATGTGTGCACCTGCAAAAAATTTGG	1695		
				Db	481	AGGCTCAATATGTCTCTTAATAGTTATGGAAAGTGTGAAAGATGTGTGCACCTGCAAAAAATTTGG	540		

Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GOOM

Center clone name: CH230-106A11

----- Summary Statistics

Assembly program: Phrap; version 0.990329
 Consensus quality: 187840 bases at least Q40
 Consensus quality: 190832 bases at least Q30
 Consensus quality: 192831 bases at least Q20
 Estimated insert size: 194114; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

* 1 207304: contig of 207304 bp in length.

Location/Qualifiers

1..207304

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-106A11"

ORIGIN

Query Match 36.7%; Score 908.2; DB 2; Length 207304;

Best Local Similarity 78.8%; Pred. No. 2.3e-149;

Matches 1122; Conservative 0; Mismatches 293; Indels 8; Gaps 3;

```

QY 199 TTGCAATCTTCAGTAAAGCAGCTTTGAATCTGAGCTTCATATCGAAGAGATG 258
DB 72806 TTTAATTTCTTAGAAGAGAGCACTTTGAATCTGAACTTGATATAAAGAGAAATG 72747

QY 259 AAAAATPACCATTTGATTAAGAAGAACTGGCTTTCTGTAGCTGGGATATCTTTTCATAGT 318
DB 72746 AAAAACCAGTTGGATTAGAAGAACTGGCTTTCTGTAGCTGGGATATCTTTTCATAGT 72687

QY 319 GTCCATCTTGGAAACATACATTTTTCAGAGCTTCGAAAGCAGCTCTGTAATAATTCAGTCT 378
DB 72686 GTCCATCTTGGAAACGTAATTAACAGAGAGCTTCGAGAGAGCTCTGTGAGGCTCAGTCC 72627

QY 379 CAAAGCAAAACAAAGAGTATTGAAGAGTGAAGTAAATAAATAATTTGGAAATTAATAATT 438
DB 72626 AGAGCAAAACAAAGGAATATTGAAGAGTGAAGCAAAATAATTAATTAATAATTAA--T 72569

QY 439 GTCAATTAATCATTTCTATGCTGATTTAGCTTCAATAACATTTGAATTTTATAGC 498
DB 72568 GTCTTTAATTCATCTATGCTGTTATTTCTGAAA-----ACCACCTGATTTGTAGCT 72514

QY 499 CCAATGCTGCATATTCATACATTTTAATTCCTAAGAAATATTTTAAATGTTAAAGCGTA 558
DB 72513 GGTAATGCTGCATGTACAGACCTTTCATACCTGAGAAATAGTCTTTAATGTTCAAAATACAT 72454

QY 559 TAAATCAATAAATAGAAAATGTTGTTTACAAAATAAAGCTCTTCACTAGTTACCA 618
DB 72453 GACATTAATAAATAAGAAAACATGTGAACAACTAAAGGTGCTTTCAACATTA-CAC 72395

QY 619 CTGAAGTAAGATGTCCTCGTTTGGAGCTAAGAGCCATCATTTGTGTGAAGTGAACCACT 678
DB 72394 CTGAACGAGATGTCCTCGTTTGAAGCTAAGAGCCATCATTTGTGTGAAGTGAACCACT 72335

QY 679 GACAACGTAGAGAGTCAGACCAACATTTCTGTCTTGAAGAAAGATTTGTAACATCATGCTA 738
DB 72334 GACAAGTGAGAGTCAGGTCTACACTTTTCTGTCTTGAAGAAAGAAATCATAGCCTCATGCTA 72275

```

```

QY 739 TGGCCCTCAGTAGGCTGAAGACGCTGCACAAATGGCTTTGGAGGTAGCTGTGTACAA 798
DB 72274 TGGCCCTTCAGGAGGCTGAAGACGCTGCACAAATGGATAGGAGGTGTGTGTGTACAA 72215

QY 799 CTCACAGCTCCTCAGCTCTGCTCAGTCACCTTTTGGTGCACACATCCCATTTTAAAGATCCT 858
DB 72214 TTCACAGCTCCTCAGCCTGCTTCGAAAACCTTTGGTGCACCCATCCCATATTAAAGGTCCT 72155

QY 859 GACACCTCCATACAGATCATGTCTCAAGCTTCAAGTATGTGTGCTTATTCACAGCTAT 918
DB 72154 AACATCATCTGTGCAGATCATGTGTGCTGTCTCAAGTATGTGTGCTGCTGCTTATTCACAGCCT 72095

QY 919 TCTTGTCTGCAACCTGATTGAAAATGTTACAGAGATTTAGCTTTGACACCCACCACTGTCTAT 978
DB 72094 TCTGTCTGCAACTGATTGAAAATATTTCAGAGAAATAGTTTGCACCCACCAACTGCTAT 72035

QY 979 TAGATTAATAAACAATCTTTTGTAGTCTTTTGCATCAGTTATCTCAAGTCTGTATPACCTGTGG 1038
DB 72034 TAAATTAATAAATACTCTTTGGTCTCAGTATCAGTTATCTCAAGTCTGTAAAGCCTGTAG 71975

QY 1039 TTGTGCAATCCAGTGGACTTTAGTAGTACTCAGATCCTCTCTTGTGTGGTGGCTAGTAT 1098
DB 71974 TTGTGCAATCCAGTGGACTTTAGTAGTACTCAGATCCTCTCTCTCAACTTTGGTACACAGTAT 71915

QY 1099 ATTAAACAGTAAACCTGCTGTATGCTCACAGAAAGGAAACAGAGCATGTCAAGTCTTT 1158
DB 71914 CTTAAACAGCAACACGCTGCTGCTTACCGAAAGGAAATAGATCATATAGTATGCTTT 71855

QY 1159 GATCCTGAGAGCTTTTGTCTTACAAATTCAGAAAATCTGAAGGCCACATCATTTTATAGG 1218
DB 71854 GATTTTGAAGCTTTTGTCTTACAAATTCAGAAAAGTGCAGAAAGCGAAATGGTTTATAGG 71795

QY 1219 AAAGAGTTAATTTGATACCTTTTAAAGGTCAAAGAGTTATAGATTTCCACTGTATTTACCTGG 1278
DB 71794 GAAGAGTATAATTTGTCTTAAAGGGCCAAACAGTAAACAGATTTCTACTGTATTTACCTGG 71735

QY 1279 GATACCTATTGAAATGTGAGAGTTCAATTTAATGAGGCTATTACCTATCAAAAATCAAC 1338
DB 71734 ACTTCTTATTGAAGCATCAGAAAGTTCAATTAAGAGATATTATCCCAACCAGAGTCAAG 71675

QY 1339 TGCCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCGAGACACATTTCTGACATGGAGA 1398
DB 71674 TACCCTTAGGGTGGCACTCTTCTGTGCATCTTTATCTGAGATTTTCTTAATGCTGGAGA 71615

QY 1399 AGGAATCTGTGTGTCAGTATAGGGTTCTCTTCAAAATGCAAGTCTTTGGACAGCTGCT 1458
DB 71614 AGGAACCTTGTGTGTCAGTATAGGGTTCTCTTCAAAATGCAAGTCTTTTAGAGCAGTGTCT 71555

QY 1459 TAACTAGAGGAGCTTAATCAGTGACCACTAGATCTTGTCTGTGCCAAAAAGTTAT 1518
DB 71554 TAACTAGAGAGAGCTAGTACGAGCAATGATAGTATCTTGTCTGTGCCAAAAAGTCAAT 71495

QY 1519 ACATCCATCTTTGAAGAGCTTTCTCAATATGATCGTATTTATTCATAGACAGAAATGG 1578
DB 71494 ACACCCATCTTTGAAGAGCTTTCTCAGTGAAACATCAAAATATTGCAATGACAGAGTTGG 71435

QY 1579 AGTGACTGTGAGAACCCCTGACTTAAATGACAGAACACAG 1621
DB 71434 AGTGACTGTGAGAACCTCTGAGCAAAAGTGCAGGTAAAAAG 71392

```

RESULT 15

AC096067

LOCUS

DEFINITION

AC096067

VERSION

KEYWORDS

SOURCE

ORGANISM

AC096067

DEFINITION

AC096067

VERSION

KEYWORDS

SOURCE

ORGANISM

AC096067 259468 bp DNA linear HTG 09-NOV-2002
 Rattus norvegicus clone CH230-20F4, *** SEQUENCING IN PROGRESS ***,
 3 unordered pieces.

AC096067

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE AUTHORS	Rattus. 1 (bases 1 to 259468) Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Ayoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,P., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,J., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Gantia,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Haves,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulsegod,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangun,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mathewine,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwakoelameh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Prannkocch,C., Plopper,F., Quindexter,A., Popovic,D., Primus,E., Pu,L.-L., Fuazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabar,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL	Direct Submission Unpublished 2 (bases 1 to 259468) Worley,K.C. Direct Submission Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 259468) Rat Genome Sequencing Consortium. Direct Submission Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Nov 9, 2002 this sequence version replaced gi:22772218. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are oriented and separated by sized gaps filled with Ns to the estimated size. The sequence
REFERENCE AUTHORS TITLE JOURNAL	COMMENT On Nov 9, 2002 this sequence version replaced gi:22772218. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are oriented and separated by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GEIS
Center clone name: CH230-20F4
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 22315 bases at least Q40
Consensus quality: 228974 bases at least Q20
Consensus quality: 232210 bases at least Q20
Estimated insert size: 226685; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

* 1 113333: contig of 113333 bp in length
* 113334 113433: gap of unknown length
* 113434 258129: contig of 144696 bp in length
* 258130 258229: gap of unknown length
* 258230 259468: contig of 1239 bp in length.

----- Location/Qualifiers
1. .259468
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-20F4"
2306. .3150
/note="clone boundary
clone end:Sp6
site:ECORI
end_sequence=BH319926"
113434. .114894
/note="wgs contig"
complement(140621..141449)
/note="clone boundary
clone_end:T7"
site:ECORI
end_sequence=BH319925"
185739. .186913
/note="wgs end_extension
clone end:T7"
207447. .209977
/note="wgs end_extension
clone end:T7"
255251. .258129
/note="wgs end_extension
clone_end:T7"

Query Match 36.7%; Score 908.2; DB 2; Length 259468;
Best Local Similarity 78.8%; Pred. NO. 2.2e-149;
Matches 1122; Conservative 0; Mismatches 293; Indels 8; Gaps 3;
QY 199 TTGCATCTTCAGTTAAGAACACGCTTGAATCTGAGCTTCATATCGAAGAGATG 258
Db 116062 TTTAATTTCTTAGAAGAGAACCACTTTGAATCTGTGATATATAAAGAGAAATG 116121

QY 259 AAAAAACACAGTTGGATTAGAAAGAACTGGCTTCTGTAGCTGGGATATCTTTTCATAGT 318
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 116122 AAAAAACCCAGTTGGATTAGAAAGAACTGGCTTCTGTGGCTGGGGTGACTTTTCGTAGGC 116181
QY 319 GTCCATCTTTGGAAACATACATCTTTTTCGACAGAGTCTGCAAGCAGTCTCTAAATTTTCAGTCT 378
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 116182 GTCCATCTTTGGAACTGCTTAAATACAGAGAGTCTGCGAGAGAGTCTGTGAGGTCTCAGTCC 116241
QY 379 CAAAGCAAA CAAAGAGTATTGAAGAGTGAAGTAAATATAATTTTGGAAATTTACTAAATTT 438
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 116242 AGAGACAAACAAAGGAATATTGAAGAGTGAAGCAAAATAAATCTTGGAAATTTAA--T 116299
QY 439 GTCAATTAATCATCTTATGCTGATTAGCTTCTAAACATTTGAACCTTTTGAATTTTATAGC 498
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 116300 GTCTTTAAATTCATCTATGCTGGTTATTGTCTGAAA-----ACCACCTTGATTTTGTAGCT 116354
QY 499 CACAATGCTGCATPATTTTCATACATTTTAAATTCCTTAAAGAAATAATTTTAAATGTTAAACGCTGA 558
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 116355 GGTAAATGCTGCATGTACAGACCTTCATACCTGGAGAAATAGTCTTTTAAATGTTCAATACAT 116414
QY 559 TAATGCAATAATAGAAAAAATGTGGTTTACAAATTAATAAACCGGTCTTCACTAGTTACCAAC 618
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 116415 GACATTAATAATAAATGAAAAAACAATGTGAACAAACTTAAAGGTGTCTTCACAACTTA-CAC 116473
QY 619 CTGAAGTAGAGTCTCTCGTTTTCGAGAGCTAAGAGCCATCATTTGTGTGAAGGTGAACCACT 678
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 116474 CTGAAGCAGGATGTCTGTTTGAAGCTAAGAGCCATCATTTGTGTGAAGGTGAACCACT 116533
QY 679 GACAACTGAGAGAGTCAAGACCACTTCTGTCTGTGAAGAAATTTGTAACATCATGCTA 738
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 116534 GACAACTGAGAGAGTCAAGTCTACACTTCTGTCTGTGAAGAAATCATAGACCTCATGCTA 116593
QY 739 TGGCCCTCAGGTAGGCTGAAGCAGCTGCAACAATGGCTTTGGAGGTACGTGTGTACAAC 798
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 116594 TGGCCCTCAGGAGGCTGAAGCAGCTGCAACAATGGATTAGGAGGTGTGTGTGTACAAC 116653
QY 799 CTCACAGTCTCAGCTCTGCTCAGTCACTCTTTTGGTCAACATCCCATTTTAAAGATCCT 858
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 116654 TTCACAGTCTCAGCCCTGCTTCGAAACCTTTGGTCAACCCATCCCATATTAAAGGTCCT 116713
QY 859 GACAGCTCCATACAGAATCATGTGTCAAGCTTTCAGTGATTGTGGCTTATTTCACAGCTAT 918
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 116714 AACATCATCTGTGCAGAAATCATCGTGTGCTGCTTTCAGTGACTGGGCTTATTTCACAGCCAT 116773
QY 919 TCTTTGCTGCAACCTGATTGAAAAATGTTCAGAGATTAGGCTTGACACCCACCACTGTCTAT 978
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 116774 TCTGTGCTGCAACTGATTGAAAAATTTCAGAGAATAGGTTTGACACCCACCACTGCTAT 116833
QY 979 TAGATTAAATAAACATCTTTTGAGTCTTTTGCATCAGTTATCTCAAGTCTGATACCTGTGG 1038
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 116834 TAAATTAATAATAATACCTCTTGGGCTCTCAGTATCAGTTATCTCAAGTCTGAAGCCCTGTAG 116893
QY 1039 TTGTCGAATCCCAGTGGACTTTAGTAGTACTCAGATCCTCTTGTGTTGGTGGGTAGTAT 1098
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 116894 TTGTCGAATCCCAGTTGACTTCAGAAGTACACATACCTTCTCCTCAACTTGGTACACAGTAT 116953
QY 1099 ATTAACAAGTAAACCTGCTGTATGCTCACAGAAAGGAAACAGAGCATGTGAGTGTCTTT 1158
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 116954 CTTAACCAAGAAACCAAGCCTGCTGCTTACCAAGAAAGGAAATAGATCATAAGGTGCTTT 117013
QY 1159 GATCCTGAGAGCTTTTTCGCTTACAAATCCAGAAATGCTGAAGGCCACATCATTTTAGG 1218
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 117014 GATTTGAAAGCTTTTTCCTTACAAATCCAGAAAGTGAAGAGCGAATGGTTTTAGG 117073
QY 1219 AAAGAGTTTAAATGTATACCTTTTAAAGGTCAAGAGTTATAGATTCACCTGTATTTACCTGG 1278
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 117074 GAAGAGTATAATGTTTCCCTTAAAGGGCCAAACAAGTAAACAGATTCTACTGTATTACCTGG 117133
QY 1279 GATACCTAATGAAATGTCAGAGTTCAATTAATGAGGCTATTACCTATCAAAAAATCAAC 1338
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 117134 ACTTCTTATTGAAGCATCAGAAGTTCAAATTAAGGAGATTATTACCCACCAGAAAGTCAAG 117193
QY 1339 TGCCCTCAAGGTGGCACTCTTTTGTACAACTTTTATCCGGAGAGCACTTCTGACACTGGAGA 1398

Db 117194 TACCCCTTAGGGTGGCACTCTTCTGTGCATCTTTATCTGGAGATTTTCTAATGCTGGAGA 117253
QY 1399 AGGAACTGTGTGGTGGTCAAGTTATGGGTTTCTCTTGAAAAATGCAGTCTTTGGACCAGCTGCT 1458
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 117254 AGGAACCTTGGTGGTCCATTACCAAGTGTCCCTTGAGAATGCAGTTTATAGACAGTTGCT 117313
QY 1459 TAACTTAGGAAGGAGCTTAATCAGTGACCACTAGATCTTGTCTGTGTCGCAAAAAAGTTAT 1518
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 117314 TAACTTAGGAAGACAGCTAGTCAGTGACCATGTAGATCTTGTCTGTGCAAAAAAGTCAT 117373
QY 1519 ACATCCATCTTTGAAGCAGTTTCTCAATATGATCATCTATTATTGCCATAGACAGAATTGG 1578
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 117374 ACACCACTCTTTGAAACAGTTCTCAGTGAACTCAATTTATGCAATCGACAGAGTTGG 117433
QY 1579 AGTGACTCTGATGGAAACCCCTGACTTAAATGACAGGAACACAG 1621
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 117434 AGTGACTCTGATGGAAACCTCTGAGCAAAAGTGACAGGTAAANAAG 117476

Search completed: September 6, 2005, 14:56:42
Job time : 10594 secs